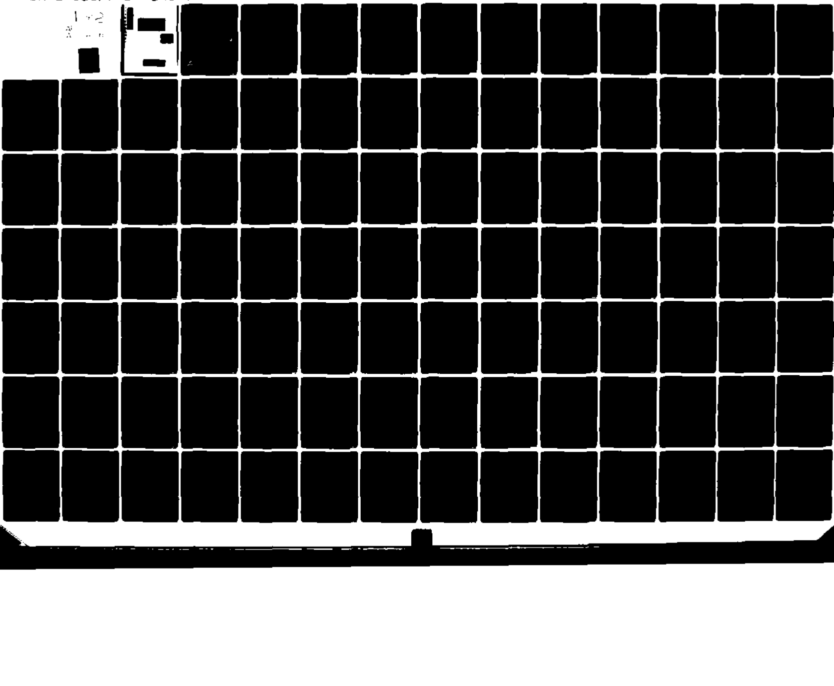
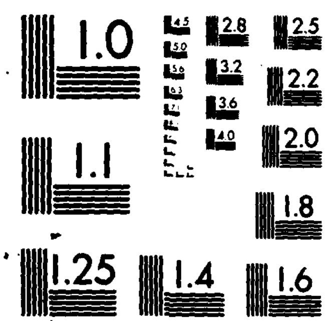
AD-A091 866 SCIENCE APPLICATIONS INC MCLEAN VA F/G 8/10

SPECIFICATIONS FOR AN STD/CTD SYSTEM AT THE NODC,(U)

JUN 80 E J MOLINELLI, R STIEGLITZ N00014-79-C-0906

UNCLASSIFIED SAI-81-195-WA NL





MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

(11)

SPECIFICATIONS FOR AN STD/CTD SYSTEM
AT THE NODC

SAI #81-195-WA

SAI
ATLANTA • ANN ARBOR • BOSTON • CHICAGO • CLEVELAND • DENVER • HUNTSVILLE • LA JOLLA
LITTLE ROCK • LOS ANGELES • SAN FRANCISCO • SANTA BARBARA • TUCSON • WASHINGTON

6

SPECIFICATIONS FOR AN STD/CTD SYSTEM AT THE NODC

14 SAI-081-195-WA

OPD TR-80-202-02

27 June 1980

Contract #N00014-79-C-0906/A-002

15

Prepared for:

E. M. Stanley
NORDA Code 500
NSTL Station, MS 39529

12 144

Prepared by:

10 E. J. Molinelli
R. Stieglitz

SCIENCE APPLICATIONS, INC.
1710 Goodridge Drive
McLean, Virginia 22102
(703) 821-4300

408404 gsm

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Section 1
INTRODUCTION

The National Oceanographic Data Center (NODC) has an obligation to archive stratification data including the modern measurements made by electronic STDs and CTDs. In a companion report (Molinelli and Kirwan, 1980) the information requirements on an STD/CTD system are determined. In this report, NODC's present system is evaluated in terms of its ability to handle the required information and in terms of the cost, speed and simplicity of the system's operation. These items are presented in Section 2. Specific improvements to the system are recommended in Section 3. Appendices are included which give more details than are presented in the text of this report.

The flow of data and information through the components of an ideal system is depicted in Figure 6.1 of Molinelli and Kirwan (1980) and is repeated here as Figure 1.1. That figure maps the relationship between the items that are discussed in this report.

The system is required to handle stratification data and station and cruise information. The data are arranged in cycles which each contain a single set of parameter values, typically, a pressure, a temperature, a conductivity and a salinity value. A quality flag value might also be included. Data cycles collected at several depths in the neighborhood of a single geographic and temporal point by identical procedures are grouped into a

STD/CTD DATA FLOW AT NODC

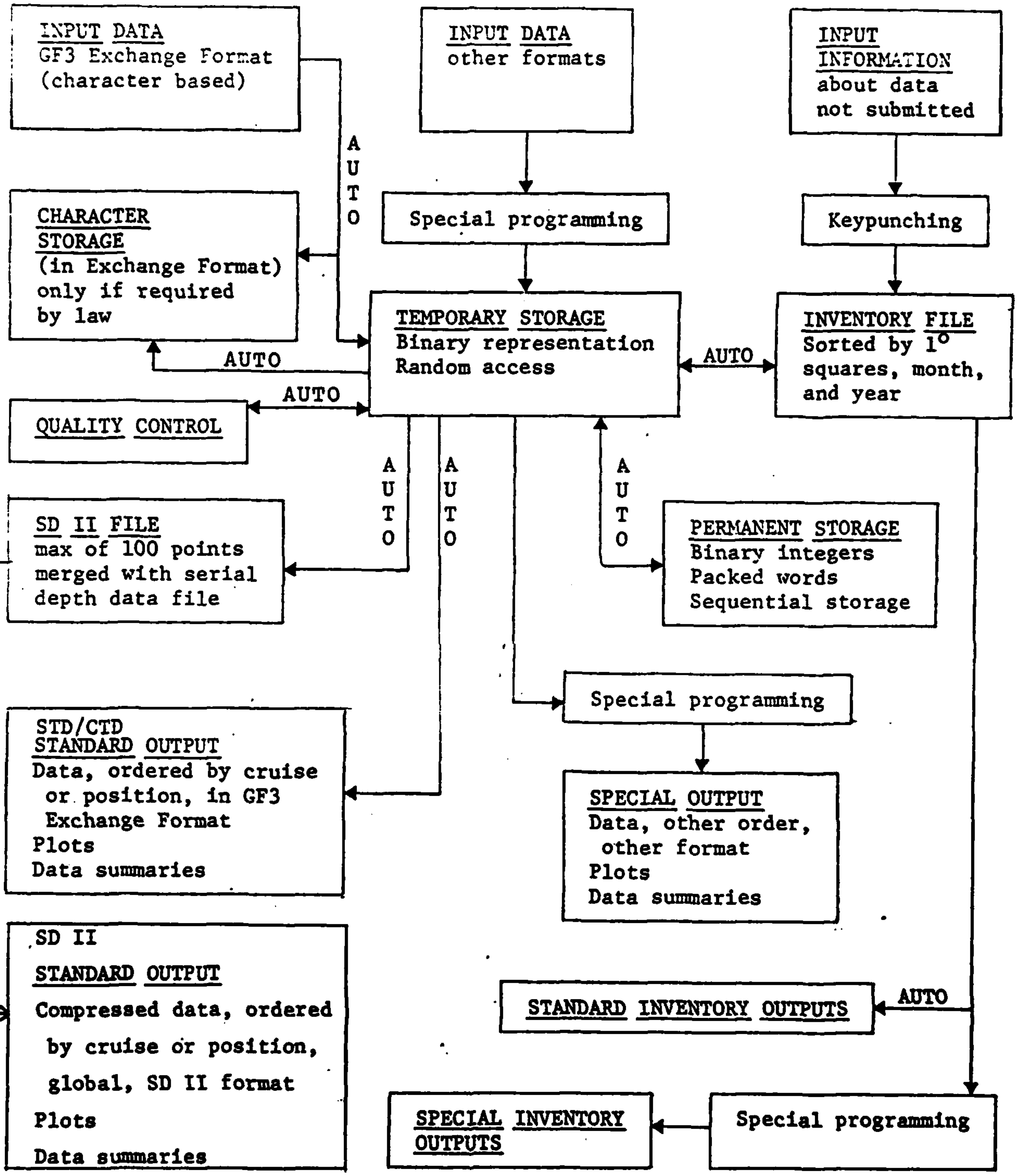


Figure 1.1

Figure 1.1 Schematic diagram of data flow through the NODC. Data flows along arrows in direction indicated assuming forms described within the boxes. Some arrows are labelled AUTO, indicating that flow along these paths can be controlled automatically by a system operator using standard programs. Only the AUTO path between the INVENTORY FILE and STANDARD INVENTORY OUTPUTS and the AUTO path between the SD II FILE and SD II STANDARD OUTPUT already exist. Some paths can only be traversed by going through "special programming" or "keypunching."

station. Information common to all data cycles at a station is reported with the station. Several stations collected consecutively over a more extended space and time range are assembled in a cruise.

Station information includes ship, cruise and station identifiers, and station time, space coordinates. Other information required by users of an STD/CTD system, as specified by Molinelli and Kirwan (1980), include: surface observations (wind, wave and weather); classical measurements (water samples of salinity, dissolved oxygen and nutrient contents, and reversing thermometer temperatures and depths) if any; collection techniques (instruments used, drop rates, sampling intervals and recording rates); parameters measured (names units, quality indicators and default values); and processing techniques (lag corrections, edits, filters, derivations, calibrations and interpolations).

According to Molinelli and Kirwan (1980) user needs are met by an archive in which data are stored on tape in cruise order, as efficiently as possible (efficiency of storage is defined later in the present report). Quality control tests are to include the determination and flagging of density instabilities, noise level and vertical resolution tests and, optionally, comparisons with climatic mean values.

Data are also to be degraded in vertical resolution and incorporated in the serial depth data file.

Section 2
THE EVALUATION OF NODC's PRESENT SYSTEM

The present system is taken to be that described by Hadsell (1974).

2.1 GENERAL MEASURES OF A SYSTEM

There are three basic measures of a computer system:

a) Can it handle the job?

b) Is it practical and/or efficient on the computer?

c) Is it easy to use?

2.1.1 *Can It Handle the Job?*

The system must be able to accomodate all parameters of interest in the given area. A complete parameter specification includes not only value, but units, precision indications and even some sort of quality control indicator. Provision should be made for a 'not measured' status value (i.e., a dummy value) in order to keep data structure simple.

An important part of any data set is some associated comments - things that cannot readily be quantified but are best left as text. Such things might include experimentor's name, institution name, processing methods, simple algorithms, etc.

It always seems despite much forethought, that something new is desired in a system. Additional parameters, more comments, differing accuracy etc. are such examples. The system must be flexible enough so that upgrades can be made with only a reasonable effort.

A final entry in this category could be called 'responsiveness'-- does a system provide the user with the information needed.

2.1.2 Is It Practical and/or Efficient on the Computer?

The relevant considerations in this category are the following, inter-related, items:

a) Response time

b) Amount of core used

c) Running time

d) Efficient use of other computer resources

e) Cost

Some of the above factors are heavily influenced by computer center considerations totally external to the program under examination. However, they must certainly be addressed in the context of a given program.

2.1.3 Is It Easy to Use?

It is absolutely necessary when designing a system or program to consider the community of prospective

users. If a code is to be used by computer scientists, then things can be organized in one way. However, in another set of circumstances, one should not be required to have the knowledge of a computer scientist to run things. This is especially true for the routine use of a system. One must make the user feel that a system is trying to help rather than that the system is simply one roadblock after another. The above remarks apply equally to the areas of job control language requirements, input preparation and output formatting.

2.2 MEASURES OF THE NODC SYSTEM

In this section, we will respond to the evaluation criteria just described for the present NODC system.

2.2.1 Can It Handle the Job?

The system presently in use allows a flexible and nearly complete description of the parameters being stored. The value is provided along with units, a precision indicator and a quality control indicator. The system is incomplete in that dummy data values are not defined. This forces either the calculation and insertion of an interpolated value or the deletion of a data level. The system is inadequate in allowing only 800 characters of comments. More comments are needed to handle the information requirements specified by Molinelli and Kirwan (1980). The limit of ten parameters for a data cycle is reasonable but unnecessarily inflexible.

Finally the system in existence in unresponsive to a certain degree. It does not require and therefore does not encourage, the storing of information necessary for the intelligent use of STD/CTD data. It relies too heavily on data supplier's perceptions of secondary user needs and consequently generally fails to provide necessary details of the collection and processing procedures that operated on the data. This lack results in the improper use of the data, i.e., the data are used in the incorrect context. The situation could be improved by recommending an exchange format that requires the necessary information.

The system is unresponsive also because it does not routinely channel a degraded vertical resolution version of the data to the serial depth data file. Most of the investigators presently clamoring for secondary STD/CTD data want a degraded version in the serial depth file.

2.2.2 Is It Practical and/or Efficient on the Computer?

The present system consists of modular programs executed in a batch mode. This approach does not have a good response time compared to the more modern interactive time-sharing mode generally available. The operator of the system may typically wait for a batch job step to run overnight.

While wasteful of sensible time and therfore personnel costs. the batch mode does still enjoy some efficiency. The amount of time required of the central processing unit (CPU) is less (i.e., a separate CPU is not required to do the bookkeeping associated with time sharing), and what CPU time is required is used at off peak hours of the day. For both these reasons the computer cost per job is less.

It is impossible to predict the savings in computer dollars or the cost in personnel dollars involved in the use of the batch mode over a yet non-existent interactive system. However, experience has shown that typically, when operator decisions are necessary in the logical flow of computer programs, the time sharing mode is more efficient. In this case, the more rapid return of results is a further bonus of the interactive mode. An STD/CTD system does require operator involvement - a fact recognized in the present system by the use of program modules.

Other inefficiencies of the present system are its reliance on line printer plots and massive data rewrites in the data editing steps. On line random access storage of data, while expensive to maintain for long periods of time, does allow corrections to be entered in the middle of a file without rewriting all the unchanged records in the entire file. As a temporary storage device in the data editing step, random access disks are much preferred to the tapes presently used. As for line printer plots, they are sometimes advantageous over higher quality flat bed and drum plots because they can be produced much faster and cheaper (and, typically, with less job control language). However, they have such poor resolution that they are difficult to use (and waste personnel time). With the availability of high speed cathode ray tube (CRT) graphics, high speed and relatively high resolution plots can be produced. CRT graphics are especially appropriate with time sharing operations. What is more, the graphic terminals required for their use are already available at NODC.

This brings up the point of hardware requirements for an interactive system. The UNIVAC computer, to which NODC must adjust, is located offsite. Consequently terminal interaction will be required even to submit jobs in the batch mode. Thus there must be some investment in terminal hardware in any case. This hardware is, in general, applicable to time sharing operations.

Another point to be made here is storage space for the archived data. The 022 format presently used, makes some attempts to minimize wasted bits, i.e., bits set aside for a parameter but not needed to represent the value of that parameter. More savings in bits can be realized by applying packing and delta format techniques. These techniques and the savings associated with them are discussed in Section 3.

Efficient data retrieval requires the rapid identification and location of data, without reading through the entire data holdings. This requirement is met by maintaining an inventory file consisting only of information necessary for a search. The information should include location and date of the STD/CTD stations as well as country, institution, ship, cruise and station identifiers. NODC already maintains an appropriate inventory, the Data Inventory Products (DIP) system, which can identify those stations (by a unique reference number) which satisfy a particular search. The DIP system should be utilized in processing requests for data. What the DIP system lacks is a means of locating the tape and file on which the station of interest resides. A second, smaller inventory is

therefore required which can take the unique reference number provided by DIP and give the tape number, file number and record numbers where the data from that station reside in the archive.

2.2.3 Is It Easy to Use?

For a user with programming background the design concepts of the present system are readily understood, but for users without, the methods appear inscrutably complex. For all users the system is bulky. It relies heavily upon the assembling of job control language (JCL) codes, and puts a great deal of responsibility on the operator for tracking various intermediate forms of the data. All job routing and edits are entered by computer cards, characteristic of older systems, but now recognized as too tedious and inconvenient to make proper use of personnel time and effort. As mentioned previously, the reliance on line printer plots also places a heavy burden on the person operating the system.

The difficulties can be alleviated by establishing a terminal based interactive system with a "freindly" preprocessor. More details on this type of processor will be given in Section 3. Here it is enough to describe it as a program that prompts the operator for the necessary information and then assembles automatically the modules and JCL required to perform the indicated function(s). The program so assembled may be run in batch or interactive mode, depending upon the input required from the operator. For operators without a programming background the friendly preprocessor is essential. For all operators, the friendly proprocessor makes the system easier to use.

The present system is also difficult to use because, by not specifying an exchange format, it is forced to handle a great number of different formats during data submission. This requires a continued high level of programming effort which is not only costly for NODC but slows down the availability of the data considerably. While recommending an exchange format will not eliminate the need for translating different supplier formats, it could reduce it substantially.

Section 3
SPECIFIC IMPROVEMENTS TO THE SYSTEM

This section looks in more detail at the improvements suggested during the evaluation of NODC's present system in the previous section. The elements discussed here fit into the overall ideal system according to the map provided in Figure 1.1.

The improvements described begin with the exchange format because that specifies all the data and information that the system must handle. The next subsection deals with the storage of data within NODC and the advantages of both permanent and temporary formats. The role of the inventory in data retrieval is included in this subsection, as is the creation of data products. The next subsection addresses the friendly preprocessor which provides the means for the operator to manipulate the data along the paths marked "AUTO" on Figure 1.1.

The final two subsections deal with the algorithms required for quality control and compression, respectively. Here compression refers to the degradation of the vertical resolution.

3.1 THE EXCHANGE FORMAT

Two reasons have been given for NODC to recommend an exchange format. One is to ensure that the proper information is supplied to NODC. The other is to enable

NODC to accept, store and disseminate the data more automatically and therefore more rapidly. The GF3 system for data exchange has been recommended (Molinelli and Kirwan, 1980). This format enjoys the further advantage of being acceptable for exchange between data centers.

A complete manual for the creation of STD/CTD data exchange tapes using the GF3 system is provided with this report as Appendix A. The reader should refer to that appendix for details. Here we merely indicate those characteristics of general interest.

The format assembles data cycles into stations and stations into cruises. One cruise is written per file. The size of a record is fixed, but the number of data records per station and stations per file is variable. The number of files (cruises) per tape is also variable.

At the beginning of the tape, information is recorded concerning the institution providing the tape and the date the tape is written as well as the computer and character set used. A list of the cruises contained on the tape is also provided.

At the beginning of each file on the tape information is recorded regarding the cruise in the file, including: the oceanographic project, if any, of which the cruise is part; the country, institution and ship; cruise name; date of the present version of the cruise data; dates of the cruise; latitude and longitude limits of the cruise; type of data collected; and the number of stations in the cruise.

At the beginning of each station in the file there is a great deal of information recorded. As with the cruise heading, the names of the oceanographic project, country, institution, ship and cruise are stored as are the date and times of the station and the latitude and longitude of the station. Surface and meteorological observations are also recorded. They include: water color and transparency; wave direction and period; sea state or wave height; wind direction; wind force or speed; barometric pressure; wet and dry bulb temperatures; weather; cloud type and cloud amount. Classical water sample measurements are recorded at the beginning of the station also. They include: pressure or depth; temperature; salinity; dissolved oxygen, silicates, phosphates, phosphorous, nitrates, nitrites and pH for several levels. Up to 35 levels of classical measurements can be accommodated. A code indicates how these measurements are related in time and space to the STD/CTD cast.

The station heading also contains information about the STD/CTD cast including drop rates, trace reported and ship roll periods. Space is left for NODC provided estimates of vertical resolution and noise levels for the temperature and salinity measurements (see quality control algorithms for a further description of these estimates).

Relatively complete documentation on collection and processing procedures are required for the intelligent use of STD/CTD data. This information is stored in rather abbreviated form in plain language comments within each station. Information recorded here includes: name and address of responsible person and manufacturers of "fish",

digitizer and recorder. For each parameter reported at the station the following information is recorded: target accuracy and precision; sensor serial number; sampling rates; lag correction reference and results; derivation reference; editing reference and results; smoothing reference and scale; interpolation reference and frequency; calibration information; and reference and results of other comparisons. (Though much of this information is fixed for a given cruise it is all repeated for every station so that each station can be used independently of the others, as in a geographic sort. It is not unreasonable, when writing a tape for data exchange, to enter this information into the computer once per cruise and have the computer duplicate it for every station on the cruise.) For a more detailed description of the recorded documentation, see Appendix A.

It should be noted here that the references listed in this portion of the data tape must indicate documents (either paper or microfiche) available from NODC. If the article referenced is a cruise report and NODC does not have a copy of that report, then a copy must accompany the exchange tape when sent to NODC. NODC should make copies of such reports, including traces of profiles and tables of data in those reports, available to users of the data upon request.

The exchange format is flexible enough to allow many different parameters to be used (see Table 7 in Appendix A) in any combination and order. Consequently, before the station data are reported information must be recorded that indicates the parameters, units and order chosen.

After all the above station information has been provided, the data cycles are recorded. Though many parameters are allowed, three are crucial. These are pressure, temperature and salinity. Due to the multiple ways salinity can be calculated from pressure, temperature and conductivity, there is some worth in reporting the measured parameter, conductivity, in addition to the derived salinity. However, this information is redundant with the reference for the salinity derivation given at the beginning of the station.

Data submitted in this format is to be read at NODC immediately. However, some decisions are required before the data can be stored. The parameters in the data cycle must be checked by the operator and converted to the proper units (e.g., depths converted to pressures using the inverse of the algorithm reported for the derivation, temperatures in ^{o}F converted to ^{o}C, etc.). The comments must be read by the operator to determine if any other special action should be taken and to ensure that the required information has been provided. References for algorithms should be requested by NODC from suppliers who have omitted them. Meanwhile cruise and station header information are passed to the inventory (DIP system) automatically. Data set status must be passed to that system by the operator. More discussion of the DIP system is presented in Section 3.2.

In order to encourage the use of the GF3 system NODC should support it in the following two ways. First, NODC should provide FORTRAN routines that will accept

surface and classical measurements and arrays of pressure, temperature and salinity and generate a GF3 tape. This is in order to simplify the conversion to GF3 for the minimal, most basic use of STD/CTD instruments. Also, NODC should supply codes that read GF3 tapes and unloads their contents into COMMON blocks. Second, NODC should make a programmer available, if at all possible, to an institution or other data supplier and help their programmers establish a system to convert their data. This programmer to programmer communication will help avoid many of the errors that occur when instructions are passed through too many intermediaries. A few days of "troubleshooting" at the major suppliers will save NODC much work down the line.

3.2 DATA STORAGE AND RETRIEVAL

While the data are being viewed by the operator to determine what conversions need be made they are residing in the core of the computer. Only one station need be in core at a time. The other stations in the cruise(s) of interest need to be somewhere easily accessed one station at a time. This is necessary to keep the number of input/ output (I/O) operations to a minimum which is crucial to keep computer operating costs to a minimum (see Appendix B, Section 2). The ability to access stations one at a time means some random access filing system need be used. The one available on the NODC UNIVAC is a bank of disk packs with the storage capacity for about 4000 typical stations at once (see Appendix B). At 100 stations per cruise, this is enough space for 40 large cruises at once. For viewing cruises newly received this capacity is many times over what

is needed. For the limited geographic sorts to be supported by this system during dissemination the capacity for 40 cruises of randomly accessible stations is very likely to be sufficient. For the rare cases when it is not, a several step sort can accomplish the job.

To retrieve data from the disc a catalogue must be maintained that gives the disc record number (of the first record if the station requires more than one) for every station on the disc. The operator indicates the station of interest and that station is automatically loaded to core for viewing. The decisions are then made and conversions performed and the data from that station can be rewritten to the disc, overriding the previous version. When all such changes have been made to the data set it should be transferred to permanent storage.

Unlike the temporary disc storage described, the permanent storage should be as compact as possible and reside on tape. Techniques for compaction are described in Appendix B. The motivation for compaction is once again that operating costs are dominated by a large number of input/output operations. Variable length records with maximum sizes as large as possible are recommended to reduce I/O operations and yet handle very different types of records (i.e., short text, long data, etc.). Many different cruises should fit on a single tape (see Appendix B). The sequence of information and formats are illustrated in Figure B.1.

There is a fundamental requirement for the capacity to locate a particular set of stations in the archive given either limits of time and position or cruise name and station limits. An inventory file must be maintained to perform that function.

NODC identifies any station it posseses by using an "accession" number, "reference" number and "consecutive" number. The accession number identifies the data tape shipment received at NODC. The reference number identifies separately individual cruises within a shipment. The consecutive number identifies the stations or records contained in the cruise.

NODC maintains a system called Detailed Inventory Products (DIP) which can identify stations that fall within any specified one degree, or standard ten degree square. The system can further discriminate by date, if desired, automatically. A particular cruise of interest can be identified simply by checking a list of reference numbers that contains ship name, country and dates as well. The stations so identified can be plotted on a map or listed with any of the following elements: file type, accession number, country, project name, reference number, cruise name, platform code (i.e., ship code), station date, station latitude, station longitude, station number, ten degree square, one degree square and parameter codes.

Having identified the stations of interest, the task still remains of locating the tape, file and records in which the station data reside in the permanent

archive. This function is not performed by DIP. A second smaller and simpler inventory is required. It should be ordered by accession, reference and station numbers and should contain the tape number, file number, first record and last record of the identified station. Searches would still be performed using DIP, but the DIP results would then be translated to data location by means of the second inventory. This second inventory can be called the Data Location Inventory. The cost of creating and maintaining the Data Location Inventory would be minimal and its existence essential to the rapid retrieval of STD/CTD data.

Another purpose of an inventory is to direct users to the location of data not held by NODC. Specialized uses of STD and CTD instruments fall into this category of data. The names of institutions and individuals maintaining control of the data should be made available. This type of information is now available from the Environmental Data Base Directory maintained by EDIS. Information received by NODC regarding STD/CTD data not supplied to NODC, should be turned over for inclusion in the Environmental Data Base Directory.

Once the data have been located and retrieved, standard products can be created from the temporary file including plots and copies of data in either cruise or (limited) geographic order.

3.3 OPERATOR LOAD AND THE FRIENDLY SYSTEM

3.3.1 Introduction

As we have seen from Section 2, the NODC system is generally reasonable in that there is no gross computer waste (although there is definite room for improvement, as discussed). It falls farthest short as far as user is concerned. SAI has had two projects in the past years that demonstrate how a "friendly" system should be designed. There are several relevant considerations.

Depending on the nature of the tasks to be performed, the job control language can become very long and complicated. The working scientist should be spared these details which certainly do not interest him.

Good interactive programs associated with graphics terminals with their quick response time, flexible nature and helpful prompts are very much preferable to operating with punched cards and a (probably more) rigid batch system. In particular, interactive programs can free the user from all sorts of tedious formatting details.

Some segments of a complex task are suited for batch work and specifically not interactive work because of computer considerations: run time, core requirements, physical tape usage, etc.

Hard copy data plots can be a big help in the interactive analysis. It may not be feasible to produce these during an interactive session

Depending on the job, some quiet thinking away from the computer may be quite important. A scientist should have a good plan of action before addressing a computer terminal.

Almost all present computer systems have the capability for a running job to generate another job stream (including JCL) on a mass storage device and trivially ship this generated job off for batch execution.

All of the above considerations are directly relevant to the NODC problem. We make one additional assumption before proposing a solution: the existence of a robust inventory file computerized and of modest size that will know about all relevant cruises, experiments, types of measurements etc., and also where the data are stored - tape number, file number, etc. This inventory file is discussed in Section 3.2.

The basic tasks performed by this system are to interrogate the inventory file, locate relevant data (by area, cruise, etc.), recover the data from archives to a usable working form and perform automated quality control or analysis as desired. In addition the system must provide data plots and edits for the scientist/operator plus relevant summary information - results of automated tests, possible and/or probable errors or problem areas. These can be studied "off-line". Also, it must provide "friendly" means for the scientist/operator to use his judgement in modifying the files by the addition of flags on the data,

quality control comments, narrative comments, data conversions etc. Finally, the system must provide a wrap-up capability to update archive files and inventory. Such a system is illustrated in Figure 3.1. The various steps in this system are discussed in Section 3.3.2 to 3.3.5.

3.3.2 Step 1

This is an interactive preprocessor. The user will be asked what sorts of data sets he is interested in: by time, data type, spatial area, cruise name or by any of the other keys present in the inventory file. The user will also be asked what is to be done with data sets meeting the stated criteria. Several options come readily to mind.

- Retrieve for further processing.
- Provide lists and or plots.
- Provide a copy of the data for external distribution.
- Perform statistical tests on the data.
- Convert data and modify flags .

The preprocessor will then construct a file (on mass storage) which will be forwarded for batch execution. all relevant job control, including requests for physical tapes, permanent ('cataloged') file activity, attaching and execution of pre-existing processors and other resource

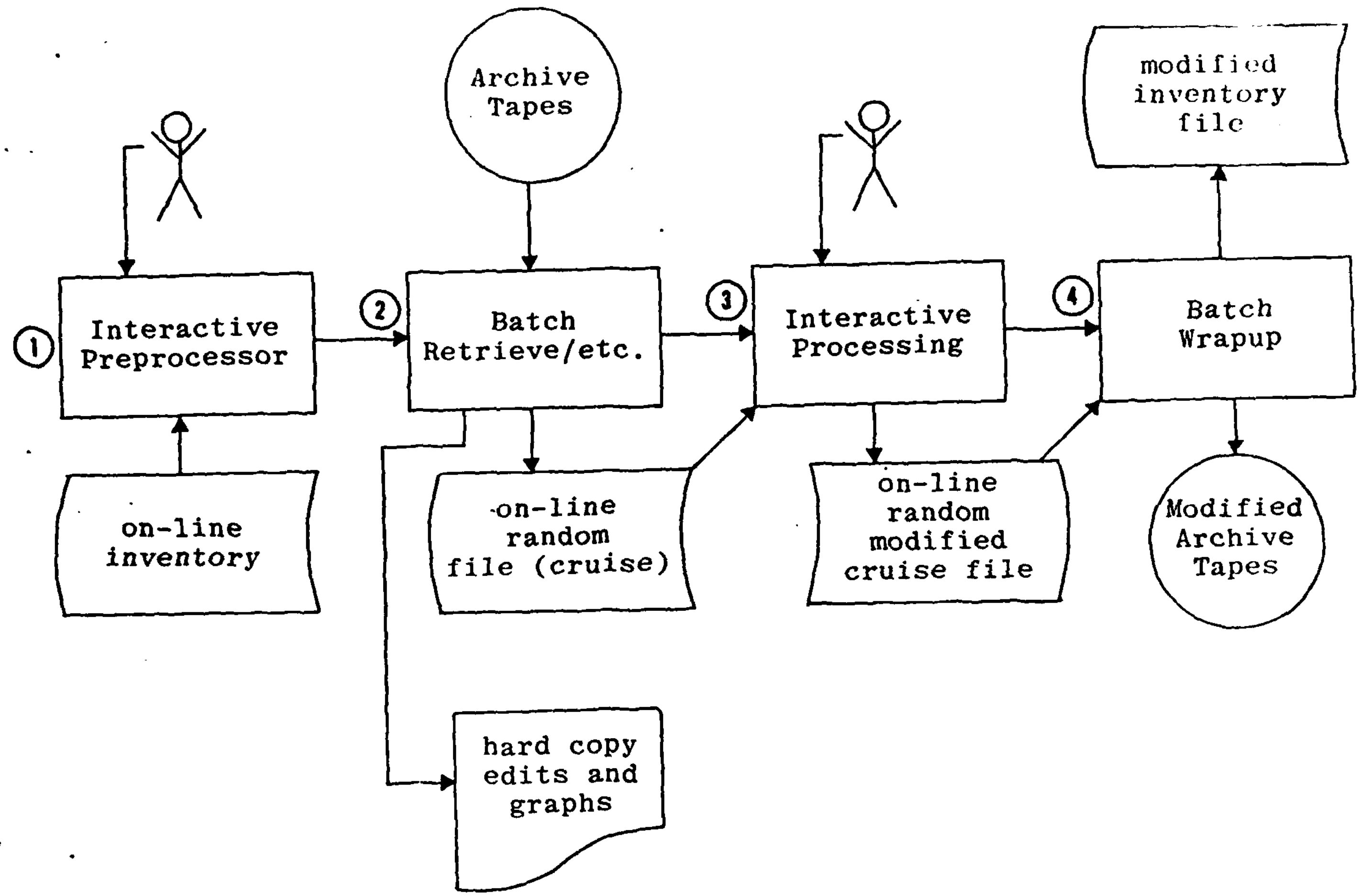


Figure 3.1 Schematic diagram of the flow of program control under a "friendly" interactive preprocessor.

requests will be automatically generated. Any necessary "data cards" will also be created. This file need never exist (unless desired) as a physical card deck since it is simply forwarded for batch execution. The scientist/ operator is almost totally freed from any purely computer-related details such as JCL generation physical tape counting, etc. which are totally peripheral to his interests.

3 3.3 Step 2

This is the job generated by the preprocessor. It encompasses all of the likely actions that should be relegated to a batch mob. Physical tape(s) will be requested and desired data sets retrieved. The data may be reformatted and saved into a random file (unpacked, from the permanent format for ease and efficieny of further use). Statistical tests which may involve some number crunching belong here, as do the generation of large edits, graphical displays and summary information. The entire process might possibly terminate with this step if the sole desired action is the generation of a data set copy (in any of several forms) for external distribution. At this point, we have the desired data sets made available for further work. The scientist/ operator is also armed with quite a bit of information (plots and results of statistical tests) about the data. He may well want to study things before proceeding.

3.3.4 Step 3

This step returns to the interactive mode. The scientist may want to modify the data sets by the

addition or deletion of data flags. addition of quality control comments. addition of further explanatory text, etc. Note that as a result of the batch job, the data file is currently stored in some random-access format which is efficient for the type of processing to be done. Output from this session might include edits and plots, on paper or on the graphics screen, and (probably) a modified copy of the internal working version of the data set. This interactive job will also generate a file for batch execution, as did step 1, when needed.

3.3.5 Step 4

This batch job (generated by Step 3) will take care of clean-up details most appropriately handled in a batch job. These would include down-loading (packing and compression) the data file for efficient permanent storage, updating archive tapes and updating the inventory file.

3.3.6 Summary

This system of codes is totally responsive to the user. He is freed from computer demands or knowledge that are irrelevant to him. The job is broken down into logical steps which are handled in the most efficient place and manner for the specific steps - either interactive or batch

The user can be provided with supplemental information before starting the second interactive step (Step 3). Thus, a good "battle plan" for that step can be devised.

It should not take a large effort to produce an original minimal system that is responsive to the user. As long as some reasonable care is exercised in the original coding, new capabilities may readily be added at a later date.

3.4 QUALITY CONTROL ALGORITHMS

By means of the friendly system described above it can be assumed that the data from one or more stations resides in core while other stations from the same cruise, at least. are randomly accessible on disc. The station(s) in the computer memory must be quality controlled. The procedures for Quality Control recommended by Molinelli and Kirwan (1980) include a test against local standards, a test for density instabilities, and tests for noise level and vertical resolution of the temperature and salinity profiles. Optionally, a calibration check against precise models of deep water potential temperature, and salinity curves is suggested. This section gives a more detailed description of the algorithms to be used to perform the necessary quality inspections.

3.4.1 Test Against Environmental Models

Available at NODC are models of salinity as a function of density (σ_t) in areas of 5° latitude by 5° longitude (D. Hamilton, NODC. personal communication).

These models are rather gross because of the great variability in the natural waters they attempt to describe. Consequentially data are grouped in large bins of salinity (.1°/00) and σ_t (.05 to .10). Such groupings can certainly not test the calibration of STDs or CTDs but they are capable of discerning some spurious values that the originators might have missed. They are ideally suited for the system described herein for they have been designed to be used with an interactive graphics terminal In this use, plots are drawn of observed pairs of S, σ_t values against a model with an envelope of acceptable salinity values as a function of σ_t for the appropriate location. Spurious values of salinity are most obvious to the operator in this mode (see Figure 3.2). He may then indicate those points whose values are believed spurious and have the quality control program flag them as questionable (see Table A.8 in Appendix A for a system of flag codes). Note that the left most observation in Figure 3.2 should not be considered spurious as it is in line with the following points.

Errors in temperature are more subtle as they may alter the σ_t value without bringing the point out of the salinity envelope. However, if temperature is far enough in error density inversions will result. These temperature errors should be apparent from the stability test described next.

3.4.2 Test for Density Instabilities

This test should be applied on a point by point basis. Each point deeper than a previous point whose local density is less than the previous point (density calculated

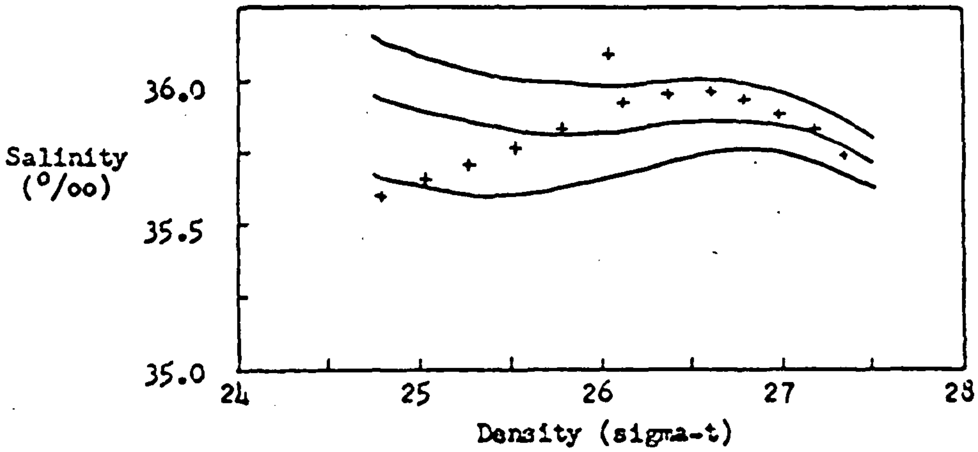


Figure 3.2 Interactive graphics display of salinity-density model (lines) and observed data (crosses). Figure provided by D. Hamilton of NODC.

at the same local pressure for both points) causes a failure of the test. Both the deeper and the shallower point are in question but further inspection is required to determine which is in error. NODC is only to flag such instabilities and is not to make corrections. The shallower point should be flagged as leading to a density instability by virtue of having too great a density (flag code 100, see Table A.8) and the deeper by virtue of having too small a density (flag code 200, see Table A.8). If the next deeper point is also less dense then what is now the middle point it also is flagged as a minimum. In this comparison the middle point is the maximum, so it is flagged again by the addition of flag 100. (It now has a flag value of 300 indicating it is unstable both with respect to the point above it and the point below it.)

For performing this test the temperature (T1, T2) and salinity (S1, S2) values of both points being compared are required as are the pressures (P1, P2). The numeral 1 refers to the shallower point.

The local density for the comparison is then determined by the following steps:

- Determine the average pressure, PA = 1/2 (P1 + P2).

- Compute the adiabatic change in temperature to bring point 1 to the pressure PA. The adiabatic

gradient for water as a function of (in situ) temperature, salinity and pressure is calculated using the formula of Bryden (1973). This gradient, G1, is then multiplied by the pressure difference (PA - P1) to give the temperature change $\Delta T1$.

- Similarly compute $\Delta T2 = G1 \times (PA - P2)$.

- Calculate the in situ density of water 1 at pressure PA. This is a function of $T1 + \Delta T1$, S1 and PA. The equation presently accepted as standard is the Ekman (1908) formula although a new set of equations is presently being developed. The Ekman formula gives a value of the specific volume which is the reciprocal of the density.

- Similarly calculate the in situ density of a water parcel with temperature = $T2 + \Delta T2$, salinity = S2 and pressure = PA.

- Compare these two values to determine if point 1 is denser than point 2. To be considered significant the difference must be greater than the effect of measurement error on the calculation. An appropriate tolerence would be .01 $\sigma_{s,t,p}$ units.

The station's T and S profiles and a profile of $\Delta\sigma_{s,t\ p}$ should then be plotted on an interactive graphics terminal with points leading to instabilities marked This will enable the operator to determine whether the program is proceeding in a reasonable fashion and whether there might be some simple correction for a great number of instabilities (e.g., a block of data being reported in order of decreasing pressure!). Experience has shown that it is unwise to process data in any way without being able to view them along the way.

3.4.3 Test for Noise Level and Vertical Resolution

This test determines the vertical wavenumber energy spectra of the temperature and salinity profiles in order to select those points at which the spectra flatten. The vertical wavenumber at the selected point indicates the effective vertical resolution The energy level of the selected point represents the energy level of the noise in the flat portion of the spectra. This level integrated over the wavenumber band gives an estimate of the total variance of the noise of the temperature or salinity measurements.

This test requires operator involvement. The system should present the spectrum on an interactive graphics terminal (as a log vs. log plot) and let the operator grapically select the point at which leveling commences. If no flattening is observed then no estimate of the noise level can be made and the effective vertical resolution is the provided vertical resolution.

This test is not performed on a station by station basis, but on a cruise by cruise basis.

- The cruise should be searched for several (5 to 10) of the deepest, most continuous stations that share a common set of sensors (i.e., the reported sensor serial numbers for the temperature, salinity and pressure measurements should not vary among the stations selected).

- The greatest common depth interval among the stations should be determined. Missing values should have interpolated values inserted in their place so that a uniform pressure series results.

- First difference (prewhiten) the points along the profile to prevent the overwhelming of spectral estimates by the high energy low wavenumber (long wavelength) components. If X_i is the ith element of the original profile and X_i' of the prewhitened profile, then define $X_i' = X_{i+1} - X_i$.

- To avoid contamination of high wavenumber components apply a cosine taper to the first and last 10% of the record. The cosine taper starts the profile at zero and brings it to full value over half a sinusoid in the first 10% of the record. In the last 10% of the record the cosine taper brings the profile from its full value to zero, again over half a sinusoid. This reduces the total energy in the spectrum to be produced by 10%.

- Generate a spectrum of the profile using fast Fourier Transform (FFT) techniques.

- Recolor the spectrum using a factor (F) that is a function of the vertical wavenumber bin.

  $$F = \frac{1}{4 \sin^2 \left(\frac{\pi k}{n} \right)}$$

  where n = the number of values in the input series (a power of 2) and k is an index that goes from 1 to n/2 for the n/2 wavenumber bins

- Correct the total energy in the spectrum by dividing the level in each bin by 0.9. This step compensates for the effects of the cosine taper.

- Do the above procedure for each station selected. Average the spectral levels in each bin for all the stations. Display the ensemble average spectrum on a log log plot.

To aid the operator, the plot should be displayed on an interactive graphics terminal. Allow the operator several tries at fitting a straight line through the portion of the spectrum at vertical wavenumbers lower than the flat, noise dominated, portion. The operator indicates two wavenumbers between which a least squares line is fit. The line is extrapolated to high wavenumbers as a guide. The operator then selects a point on the line at which the energy level is the same as the flat part of the spectrum.

Figure 3.3 An ensemble vertical wavenumber (k) spectrum of salinity variations (F_s) at five STD stations in the northeast Atlantic on a log-log plot. $F_s(k)$ has units of $(^o/oo)^2/(\text{cycles/dbar})$ and k of cycles/dbar. A plot with a linear wavenumber axis demonstrated flattening at the high wavenumbers which is not so apparent on this log-log plot. The noise level is approximately $10^{-5}(^o/oo)^2/(\text{c/dbar})$ and occurs at a wavelength of $10^{.42} = 2.63$ decibars. The original data were reported at one decibar and .01 $^o/oo$ increments. The wavenumber band is 0 to .5 cycles/dbar and the expected noise due to the quantizing interval is .003 $^o/oo$. The observed rms noise is $\sqrt{.5\text{x}10^{-5}} \approx .003^o/oo$, in agreement with the quantizing noise. Because of the exceedingly low energy at the high wavenumbers it is suspected that these data were low pass filtered before rounding to the nearest $.01^o/oo$. If this is so there is probably more noise in the lower wavenumbers than can be estimated from the high wavenumbers and the $.003^o/oo$ rms noise is an underestimate.

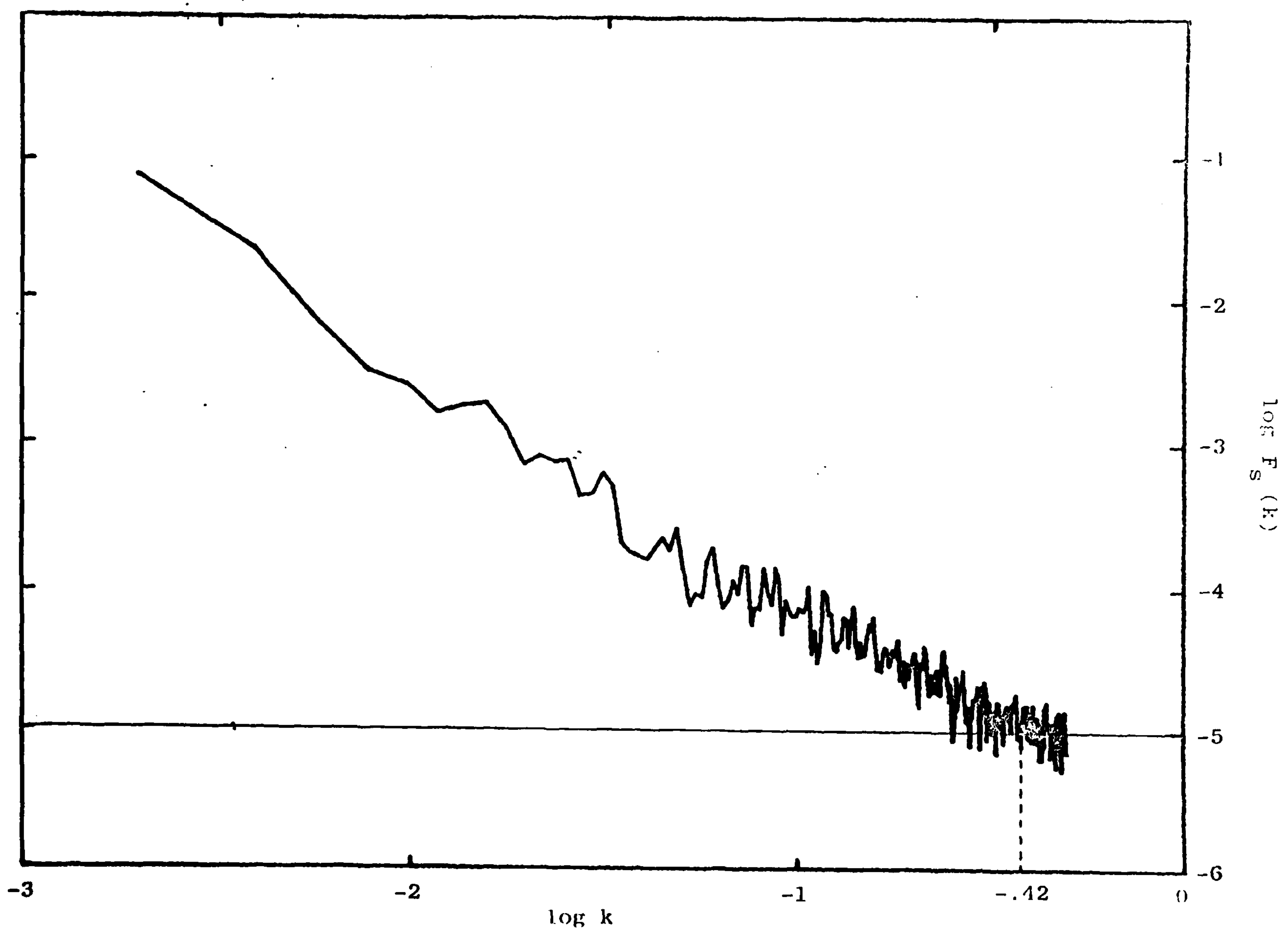


Figure 3.3 Ensemble averaged salinity spectrum.

One half the vertical wavelength associated with the vertical wavenumber of this point is stored on all stations of the cruise. It is the effective vertical resolution of that parameter (T or S) during that cruise. (Note, the station header indicates whether a particular station contributed to the ensemble average.)

The energy level of the selected point is multiplied by the high wavenumber cut off of the spectrum. The square root of this number is taken to be the rms noise of the parameter (T or S). This value is also reported for all stations on the cruise.

An example of this procedure is illustrated in Figure 3.3.

3.5 COMPRESSION

For the climatological uses of STD/CTD data a profile reduced to under 100 points and inserted in the Serial Depth Data file is normally sufficient (Molinelli and Kirwan, 1980). Standard levels no farther apart than 200 meters are desirable as are inflection points that preserve the major features of the plot. Therefore, the following compression scheme is recommended.

First, convert pressures to depths using calculations of in situ specific volumes, gravity and the equation

$$Z^* = 10^4 \int_o^{P^*} \frac{\alpha(T,S,P)}{g(P)} \, dP$$

where P^* = pressure of interest in dbars

Z^* = depth in meters at pressure P^*.

α = in situ specific volume, a function of in situ temperature (T), salinity (S) and pressure (P).

$g(P)$ = gravitational acceleration, a function of latitude and depth. For calculation of g assume depth in meters = pressure in dbars.

Second, determine values of all parameters (except pressure) at the extended standard levels (see Table 3.1). These levels come from recommendations made by A. F. Amos (personal communication) modified by increasing the maximum increment from 100 m to 200 m. Interpolate when necessary. If the standard levels are within 2 m of the observations consider the standard level values observed, otherwise mark them interpolated. Use the same 3-point Lagrange interpolation scheme presently in force in the Serial Depth Data System, with the same exceptions.

After the standard values have been selected then test the original profile against linear fits between the standard levels. As with the ICES compression, if the difference exceeds a given tolerence, say .02°C and/or .02 °/oo, add a level at the depth of maximum difference. Refit lines between the new point and the adjacent standard levels and continue comparing against the original profile.

Table 3.1

101 EXTENDED STANDARD LEVELS (meters)

Increment:	2	5	10	20	25	50	100	200	200
	0	35	60	120	225	550	1100	2200	6200
	2	40	70	140	250	600	1200	2400	6400
	4	45	80	160	275	650	1300	2600	6600
	6	50	90	180	300	700	1400	2800	6800
	8		100	200	325	750	1500	3000	7000
	10				350	800	1600	3200	7200
	12				375	850	1700	3400	7400
	14				400	900	1800	3600	7600
	16				425	950	1900	3800	7800
	18				450	1000	2000	4000	8000
	20				475			4200	8200
	22				500			4400	8400
	24							4600	8600
	26							4800	8800
	28							5000	9000
	30							5200	9200
								5400	9400
								5600	9600
								5800	9800
								6000	

This method should keep the number of points under 100 except for very variable and very deep profiles. It constitutes an improvement over the present version by eliminating the occurrence of an inflection point so close to a standard level as to be redundant. It also keeps the routine rapid and makes the inflection points much less sensitive to the starting point of the search.

Write the compressed station on a tape for reading by the Serial Depth Data System. Though frequent occurrence of compressed stations with more than 100 points is bulky, up to 160 levels can be accommodated (I. Perlroth, NODC, personal communication). If more levels than this are selected, the inflection point criteria will have to be slackened.

3.6 CONCLUSION

An outline has been presented which specifies the components that belong in a system to archive, control and disseminate historical stratification data using STDs and CTDs. Further descriptions can only be generated during the programming necessary to implement the system. It is envisioned that improvements may still be made to the components described in this report. These improvements, like the further descriptions, are appropriate during the implementation phase. It is not envisioned that any substantial changes to the system described herein are required.

Oceanographic community impatience for an historical STD/CTD archive dictates that a decision on the suitability of this system and its subsequent implementation be undertaken rapidly.

REFERENCES

Bryden, H. L., 1973: New polynomials for thermal expansion, adiabatic temperature gradient and potential temperature of sea water, Deep Sea Res., 20, 401 - 408.

Eckman, V. W., 1908: Die Zusammendruckbarkeit des Meerwassers. Conseil Perm. Intern. p. l'Expl. de la Mer (ICES), Pub. de Circonstance, No. 43, 47 pp.

Hadsell, P. R., 1974: Report of the NODC STD Workshop, the National Oceanographic Data Center, EDIS, NOAA, Department of Commerce, Washington, D.C.

Molinelli, E., and A. D. Kirwan, 1980: Requirements for an historical stratification file using STD and CTD data, SAI-81-179-WA, Science Applications, Inc., McLean, VA.

Appendix A
GF3 Exchange Format for
STD/CTD Data

A.1 INTRODUCTION

This manual for the implementation of the GF3 system for STD/CTD data exchange relies heavily on the original GF3 manual (Intergovernmental Oceanographic Commission, 1979) and will repeat entire sections. This is done so that the present writing can stand alone. STD/CTD specific paragraphs and tables will be marked by a "$".

The IOC General Format 3 (GF3) is a system for formatting oceanographic data onto magnetic tape for international exchange among data centers. Although it has not been intended or recommended for internal use in the centers, it may nevertheless be found suitable for archival of certain data, e.g., project data sets. In fact, in spite of its apparent complexity, it can be easily used by single scientific users, small institutions, or large data handling centers. Testing has shown it to be sufficiently general to encode virtually any type of digital oceanographic data including physical, chemical, biological, geological, meteorological and geophysical data. Being multi-disciplinary in nature, it also may well find use in environmental studies outside oceanography. Data can be written or retrieved with simple, rather short programs, and fully automated if so desired.

GF3 is not a wholly-new system; rather, it evolved from the format originally developed for use in the GATE project, and that was subsequently modified by IODE to become the GF2 format. GF2 suffered from three main problems: (a) a structure that was too rigid and limited to cover all types of oceanographic data effectively, (b) a

restricted capacity for the inclusion of plain language documentation, and (c) an excess of fixed format fields that were either rarely used or could just as effectively have been included as plain language comments or user-defined fields. However, as GF2 had a certain degree of acceptance and was in use at some institutes and data centers, GF3 was built from its framework. Indeed, it is possible to create GF3 files that simulate GF2 files, and files can be converted from GF2 to GF3 with minimal loss of information.

It has been emphasized in the preceding paragraphs that GF3 is a system rather than strictly a format in the conventional sense. As a system it allows the user a number of ways of organizing his data such that whichever way he chooses, the data fit into GF3. Furthermore, provision is made for the user to describe, on the same magnetic tape that carries the data, the exact format he has selected and all codes he has used, as well as ample space for plain language documentation.

This flexibility is obtained from the three major attributes of the system. First, GF3 recognizes that most data sets are organized in a hierarchy, i.e. the volume of information increases with depth in the hierarchy. Likewise, the information at the top pertains to the whole set, and information in middle levels serves to group variously together what falls below it. In GF3 terminology, the bottom of the hierarchy comprises data cycles; the middle levels group the data cycles into series; and at the top, the series are grouped into files. $ For STD/CTD applications the data cycle contains the observations made at a single level; the GF3 series contains all the data pertinent to a single station; and the GF3 file contains all the stations in a single cruise.

Secondly, GF3 recognizes that the need for format uniformity is greatest at the top of the hierarchy and that flexibility must increase in descent. Consequently, there are several mandatory fields at the file level, and at the series level, and virtually none below. Thus, the user preparing data for GF3 is given flexibility in organizing and writing his data, and the user reading data in GF3 can expect the particular high-level information he needs to understand the data he has received. $ For STD/CTD applications the user retains flexibility only in the number and types of parameters that constitute a data cyle. Station information must be supplied in a given format.

Thirdly, GF3 recognizes that its magnetic tapes must be easily read by relatively simple computer programs. At each hierarchy level in the data set, GF3 provides for information to be recorded that explains the formats the user has chosen for the descendant levels. This facility, combined with the liberal use of plain language text, means that a GF3 data set on magnetic tape is both self-describing and self-documenting. The user reading the data tape need only know that it is in the GF3 system. More simple, less general, programs can be used to read the STD/CTD implementation of the GF3 tapes.

A.2 STRUCTURAL SPECIFICATION

The best understanding of the GF3 system is gained by examining the overall tape features in terms of files, followed by an explanation of the build-up of the files themselves from the allowed record types. Finally the design of the record types will be discussed. Detailed layout of the fields within the records is found in the next section.

A.2.1 Magnetic Tape Physical Specifications

A.2.1.1 GF3 is a character oriented format for use with digital magnetic tapes. Such tapes should be 0.5 inch (12.7 mm) wide with a maximum reel diameter of 10.5 inches (266.7 mm) and a maximum tape length of 2,400 feet (732 m).

A.2.1.2 Unless agreed otherwise between exchanging parties, tapes should be either 7-track, 556 b.p.i. (bytes per inch), even parity or 9-track, 800 b.p.i., odd parity and written by Non-Return-to-Zero (NRZI) encoding.

A.2.1.3 Unless agreed otherwise between exchanging parties 7-track tapes should be written in BCD (Binary Coded Decimal) and 9-track tapes in ASCII, EBCDIC (Extended BCD) or Minsk-32 code. GF3 includes a translation table for use in the conversion of character codes.

A.2.1.4 The GF3 character set is restricted to the Latin capital letters, the decimal numerals and those special characters which are common to all four standard codes (BCD, ASCII, EBCDIC and Minsk-32). (See GF3 Code Table (2)).

A.2.1.5 All GF3 physical records (blocks) on tape are of a fixed length of 1920 bytes (characters) padded with blanks or nines (9's) where necessary to make up the length. These physical records are separated by inter-record gaps (IRG) and are organized into files. Each file is terminated by a single End of File (EOF) Mark (sometimes called a tape mark) except for the last file on each tape which is terminated by two EOF marks. Other uses of EOF marks are prohibited.

A.2.1.6 Except for the EOF mark the use of other system labels and blocks is prohibited. The EOF mark itself is a single byte block consisting of the Device Control Character, DC3 ("1" bits in tracks 2, 3 and 8 only) on 9-track tapes and the octal characters "17" on 7-track tapes. Should EOFs other than these come into use they will be added to GF3 Code Table (2).

A.2.2 GF3 Logical Tape Structure

A.2.2.1 Four types of files are recognized:

Test File
Tape Header File
Data File
Tape Terminator File

A.2.2.2 Each GF3 tape must begin with a file of Test Records terminated by an EOF mark and followed by a Tape Header File containing a Tape Header Record as its first record. Following the Tape Header File are inserted as many individual Data Files as are required before the tape itself is finally terminated by a Tape Terminator File consisting solely of an End of Tape Record followed by two EOF marks. The tape structure is diagrammed in Figure 1.

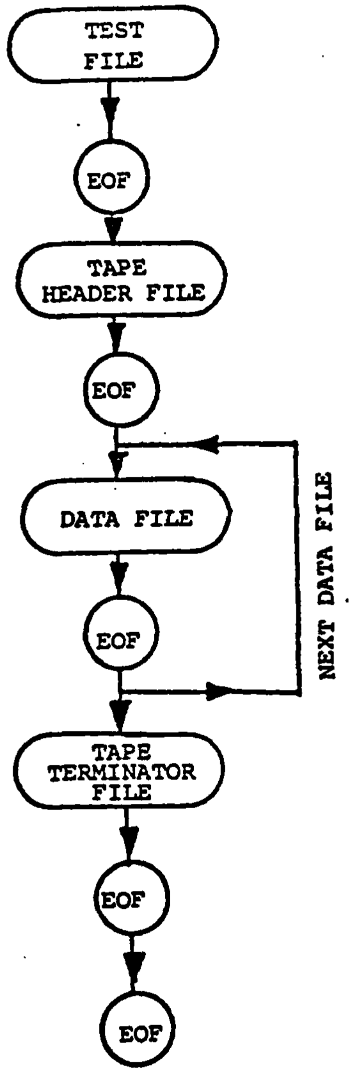


Figure A-1: GF3 TAPE STRUCTURE

A.2.2.3 If a Data File is too long to fit onto one magnetic tape, it can be continued on further reels. The Tape Header File contains information to show if the Data File is the continuation of a Data File on a preceding tape, and the Tape Terminator File shows whether the Data File is continued on a following tape.

A.2.2.4 Following the discussions of files and records in Sections 2.3 and 2.4, the organization and ordering of GF3 records within the file structure of GF3 is shown in Figure A.2 for STD/CTD tapes.

A.2.3 GF3 File Structure

Each of the four types of files identified within GF3 have a well-defined structure of allowable record types. $ The allowable types for STD/CTD files are a subset of the GF3 allowable types.

A.2.3.1 Test File

The Test File is a special file containing sufficient physical records (Test Records) to occupy about 2 meters at the beginning of the tape. Each Test Record should contain 1920 test characters defined as all binary 1's. The Test File not only provides useful protection for the beginning of the tape against mechanical damage but also allows checks on the relative alignment of tape read heads between the parties involved in data exchange.

A.2.3.2 Tape Header File

$ The Tape Header File comprises a Tape Header Record followed by Plain Language Records, and a Series

Header Definition Record. The information in the Tape Header File pertains to the entire tape.

A.2.3.3 Data File

$ The Data File comprises one File Header Record and several Series Header Records (each associated with a Data Cycle Definition Record), several Plain Language Records indicating collecting and processing procedures, and several Data Cycle Records containing the STD/CTD measurements.

A.2.3.4 Tape Terminator File

The Tape Terminator File appears only as the last file on a tape, and all GF3 tapes must end with a Tape Terminator File. This file comprises only an End-of-Tape Record.

A.2.4 $ Record Structure

All physical records consist of 1920 characters which can be divided logically into 24 cards of 80 characters each. The records are defined in general below and in more detail in a later section. The arrangement of these records on the tape within the file structure is illustrated in Figure A.2.

A.2.4.1 Test Record

The Test Record comprises 1920 characters all set to the hexadecimal "FF". This is equivalent to

GF3-TYPE TAPES FOR

STD/CTD DATA EXCHANGE

Arrangement of records for typical tape

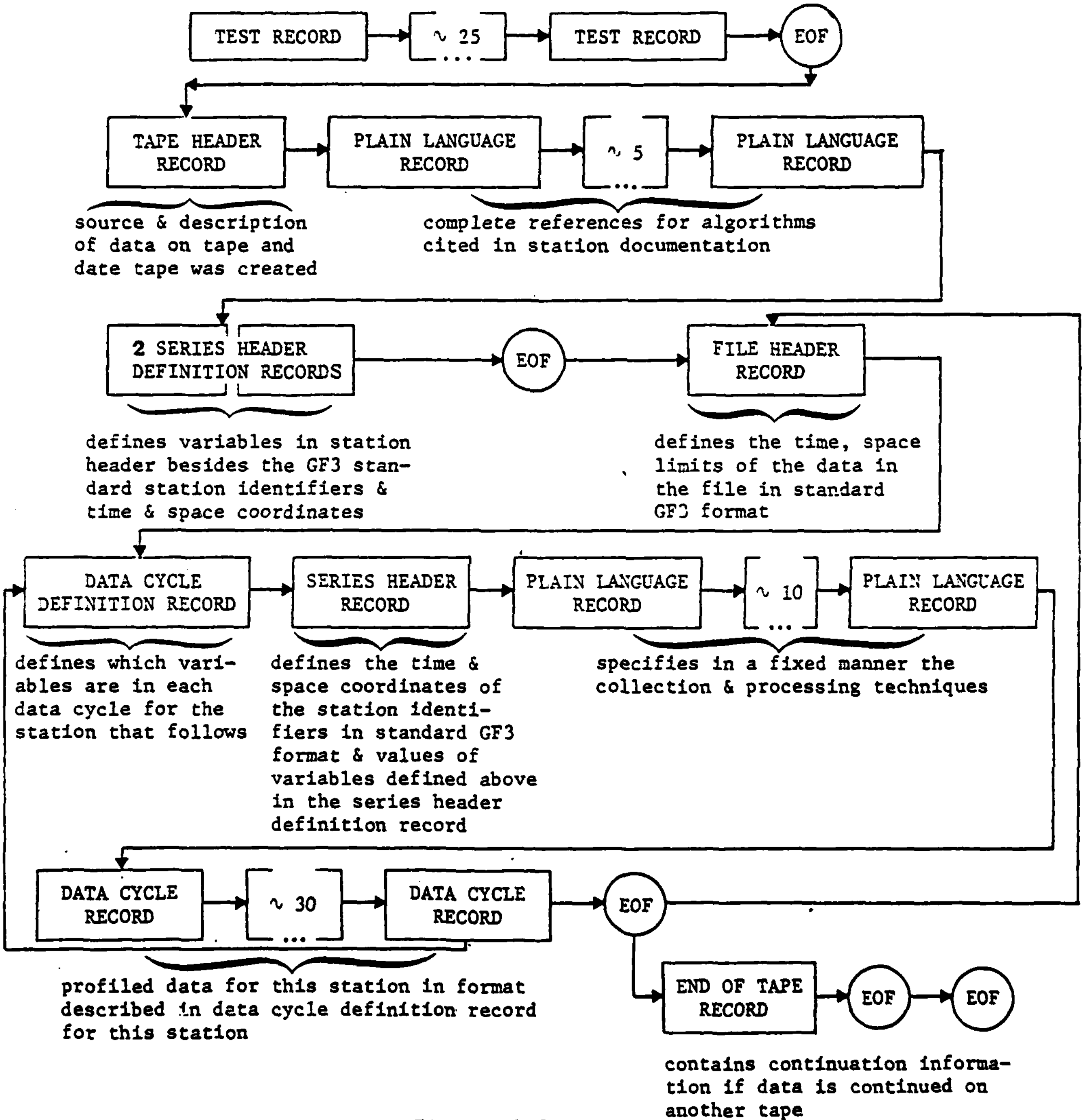


Figure A-2

setting all binary bits to "1". The Test Record is used only in the Test File at the beginning of a GF3 tape. The Test File must contain sufficient physical records to occupy about 2 meters at the beginning of the tape. The Test File provides useful protection for the beginning of the tape against mechanical damage and allows checks on the relative alignment of tape read heads between the parties involved in the data exchange.

A.2.4.2 Tape Header Record (Record ID = 1)

The Tape Header Record appears only once per tape at the beginning of a Tape Header File. The first two card images in a Tape Header Record contain information on the creation of the tape such as the tape identifier, date and time of creation, the institution responsible for its creation, and the type of computer used, together with tape sequencing information if the data reside on more than one tape. The third card image contains the standard GF3 Character Set as written by the particular computer generating the tape. This is necessary for producing a decoding table which is useful in translating characters stored with unusual codes or with unknown application of the parity bit. The remaining card images are used for plain language comments identifying the data as STD and/or CTD data and listing the cruises on the tape (one cruise per file). Other comments relating to the tape as a whole can be inserted if room allows.

A.2.4.3 Plain Language Records (Record ID = 0)

Except for the first two characters and last three characters on each card image the Plain Language Records are available for free format text in the GF3 system.

For STD/CTD data exchange the first few Plain Language Records following the Tape Header Record are reserved for listing the complete references for algorithms cited in the data processing documentation portion of the station data. Other comments on the tape as a whole may be made in following Plain Language Records.

For STD/CTD data exchange Plain Language Records are also used following the header record for each station. In this use the first several Plain Language Records are not free format but must contain specific items of information in specific positions. These items include descriptions of pertinent collection and processing procedures that apply for that particular station. (Although these collection and processing procedures are generally applicable for an entire cruise, the information is repeated so that each station may be extracted and used independently).

A.2.4.4 Series Header Definition Record (Record ID = 3)

Each STD/CTD station constitutes one "series" in the GF3 terminology. Each file is made up of several stations linked by a common cruise or location or time. For data submission only groupings by cruise are acceptable.

The Series Header Definition Record defines the contents of those card images in the Series (station) Header Record that are not determined by the GF3 standard formats. These contents have been fixed for STD/CTD data exchange. The Series Header Definition Record is written once per tape, as part of the Tape Header File, because all station headers are to be in the same format throughout the tape. The Series Header Definition Record describes the format for storing such information as: whether the data is from an uptrace or a downtrace; water color and transparency; wave direction, period and height; sea state; wind force, direction and speed; barometric pressure, dry bulb temperature; wet bulb temperature; weather; cloud type and cloud amount codes; drop rates; ship roll; and results of quality control tests. The Series Header Definition Record also gives the parameter names and format for bottle ("classical") measurements taken at the STD/CTD station. Note, data for all these parameters are stored in the Series Header Record - there are no data in the Series Header Definition Record. The Series Header Definition Record merely contains names and formats for the data.

A.2.4.5 File Header Record (Record ID = 5)

All STD/CTD stations from a single cruise should be grouped into one file for data submission. The File Header Record contains the latitude, longitude, date and time limits for the cruise contained in the file. It also names the platform, the instruments used and the number of stations in the cruise. This information is stored in the first 5 card images of the File Header Record in a standard GF3 format. The File Header Record is the first record in the data file.

A.2.4.6 Data Cycle Definition Record (Record ID = 4)

The basic STD/CTD data cycle contains pressure, temperature and conductivity measurements, and the derived parameter: salinity. Other parameters might also be reported (e.g., time, sound velocity, water velocity etc.) so a Data Cycle Definition Record is necessary to indicate which parameters are reported. Though this information is usually constant throughout a cruise, the information is repeated so that, once again, each station can be retrieved from the cruise file independently. Flag bytes should also be used to indicate questionable or interpolated elements in a data cycle. The presence of such flag bytes must be indicated in the Data Cycle Definition Record. To repeat, there are no data in the Data Cycle Definition Record, only parameter <u>names</u> and <u>format</u>.

A.2.4.7 Series Header Record (Record ID = 6)

This record occurs once per station and contains station location, time and identification information in a standard GF3 format. it also includes non-STD/CTD data such as meteorological and surface observations and "classical" measurements, all defined in the Series Header Definition Record which is described above.

The Series Header Record is followed by several Plain Language Records which indicate the collection and processing procedures whereby the STD/CTD data were produced.

A.2.4.8 Data Cycle Record (Record ID = 7)

Many of these records follow the Plain Language Records for each station. Each Data Cycle Record contains an integral number of data cycles with parameters in a format described in the Data Cycle Definition Record at the beginning of the station. In general, these records would constitute the largest portion of the records on tape.

A.2.4.9 End of Tape Record (Record ID = 8)

This Record appears only once per tape at the end of a tape. The End of Tape Record is the only record in the last file on the tape. It indicates whether the data are continued on another tape. In addition the End of Tape Record may contain comments concerning the tape.

A.3 Detailed Specification for Record Contents

In this section, specifications are given for each type of record described in the previous section, except the Test Record which is adequately specified there. The record specifications are presented in Tabular form, and each field is described under four headings:

CARD NO. -	Sequence number within the record of the card image containing the field.
CARD BYTE NO. -	Position occupied by the field within the card image expressed in bytes (characters) relative to the beginning of the card image.
NO. OF BYTES -	Length of the field in bytes (characters).
DESCRIPTION OF FIELD AND COMMENTS -	Instructions on what information to insert in the field and the manner in which to do it. Certain information is inserted via codes. Tables of these codes are included at the end of the appendix. Reference to the Tables are made under this heading.

Data coding forms that summarize the information in the following sections are included. All numeric entries are made as integers, the placement of the decimal is either inferred or entered separately. Entries in the Plain Language Records are typically alphanumeric.

A.3.1 PLAIN LANGUAGE RECORD
RECORD ID 0

Card No.	Card Byte No.	Number of Bytes in Field	Description of Field and Comments
1	1	1	"0" Plain language record identifier (set this byte to "0").
	2	1	Record identifier of the next record.
	3	75	Plain language comments or description.
	78	3	"1" Card sequence number (set these bytes to "001").
2-24		1840	Bytes 2 to 77 on each card may be used for plain language comments or description - on each card used in this way, byte 1 contains "0" the record identifier and bytes 78 - 80 contain the card sequence number (002 - 024).

Note: Plain language comments or description may be continued on succeeding Plain Language Records if necessary using card sequence numbers 025 - 048, 049 - 072, etc.- each record used in this way should be formatted as above and the 2nd byte of the record should contain the record identifier of the next record.

Any cards in a Plain Language Record not completed should be filled with blanks except for bytes 1 and 78 - 80. For the Plain Language Records which follow the Tape Header Record, the comments must contain a complete reference list for processing and collecting algorithms used on the data.

Figure A-3

GF3 Coding Form A.

PLAIN LANGUAGE RECORD

RECORD ID | NEXT REC | PLAIN LANGUAGE COMMENTS OR DESCRIPTION | CARD SEQUENCE NUMBER

CODED BY | DATE | PAGE OF

A.3.2 TAPE HEADER RECORD
RECORD ID 1

Card No.	Card Byte No.	Number of Bytes in Field	Description of Field and Comments
1	1	1	"1" record identifier
	2	1	Record identifier of the next record.
	3	4	Unassigned - leave blank.
	7	2	IOC Country code of the country of the institution that wrote (originated) this tape. Table 1.
	9	1	Enter "9" to indicate that following code is National.
	10	3	National Institution code (if available) of the institute or data center that wrote (originated) this tape.
	13	12	Tape name or number - unique to institution that wrote (originated) this tape.
	25	5	Unassigned - leave blank.
	30	12	Name or number of preceding tape (if file continued from another tape). (Blank if not a continuation tape).
	42	18	Country - plain language - name of country of the institution that wrote (originated) this tape
	60	18	Institution - plain language - name of the institute or data center that wrote (originated) this tape.
	78	3	Card sequence number "001"
2	1	1	"1" record identifier
	2-7	6	Date (YYMMDD) that this tape was written by above institution.
	8	6	Date (YYMMDD) that first version of the data on this tape was written by above institution (same as bytes 82 - 87 unless previous versions were in error or lost, etc.).
	14	6	Date (YYMMDD) that this tape was received at receiving data center or institute (must be set to 9's when tape is written - can only be "filled-in" if receiving institution copies the tape).

A.3.2 TAPE HEADER RECORD (Continued)
RECORD ID 1

Card No.	Card Byte No.	Number of Bytes in Field	Description of Field and Comments
	20	6	Date (YYMMDD) that first version of this tape was received (must be set to 9's when tape is written - can only be "filled-in" if receiving institution copies the tape) (same as bytes 14 - 19 unless updated)
	26	12	Type of computer used to write this tape (PLAIN LANGUAGE) - Manufacturers Model
	38	5	Acronym of the format used - set these bytes to "GF3.1"
	43	35	Unassigned - leave blank
	78	3	Card sequence number "002"
3	1	1	"1" record identifier
	2	53	Translation table (Table 2) containing GF3 character set
	55	19	Unassigned - leave blank
	74	4	Record size in bytes (set these bytes to "1920")
	78	3	Card sequence number "3" (set these bytes to "003")
4-24		1680	Bytes 2 to 77 on each card may be used for plain language comments or description - on each card byte 1 contains "1" the record identifier, and bytes 78 - 80 contain the card sequence number (004 - 024)

Note: Plain language comments or description may be continued on succeeding Plain Language Records if necessary using card sequence numbers 025 - 048, 049 - 072, etc.

GF3 Coding Form B.

TAPE HEADER RECORD

REC ID (1)	NEXT REC (2)	BLANKS (3–6)	COUNTRY CODE (7–8)	CODE FLAG (9)	INSTITUTION CODE (10–12)	TAPE NAME OR NUMBER (13–24)	BLANKS (25–29)	NAME OR NUMBER OF PRECEDING TAPE (30–41)	DATA SUPPLIER - NAME OF COUNTRY (Plain Language) (42–59)	DATA SUPPLIER - NAME OF INSTITUTION (Plain Language) (60–77)	CARD SEQUENCE NUMBER (78–80)
1											001

REC ID	DATE TAPE WRITTEN: THIS TAPE (Y Y M M D D)	DATE TAPE WRITTEN: FIRST VERSION (Y Y M M D D)	DATE TAPE RECEIVED: THIS TAPE (Y Y M M D D)	DATE TAPE RECEIVED: FIRST VERSION (Y Y M M D D)	TYPE OF COMPUTER USED (Plain Language)	FORMAT ACRONYM	BLANKS	CARD NO
1						GF3-1		002

REC	TRANSLATION TABLE (b=BLANK; ¢¢=ALL 1's, OCTAL 77, HEX FF)	BLANKS	RECORD SIZE	CARD NO
1	1234567890=:>b/STUVWXYZ,(-JKLMNOPQR*];+ABCDEFGHI·)[<		1920	003

REC	PLAIN LANGUAGE COMMENTS OR DESCRIPTION	CARD NO
1		004
1		005
1		006
1		007
1		008
1		009
1		010
1		011
1		012
1		013
1		014
1		015
1		016
1		017
1		018
1		019
1		020
1		021
1		022
1		023
1		024

CODED BY ______ DATE ______ PAGE ___ OF ___

Figure A-4

A.3.3 STATION HEADER DEFINITION RECORD #1
RECORD ID 3

Card No.	Card Byte No.	Number of Bytes in Field	Description of Field and Comments
1	1	1	"3" identifies this record as a series (station) header definition record, i.e., is used to define non-GF3 areas of the station header record.
	2	1	Record ID # of next record on tape. $ It is also a station header definition record, so set to "3".
	3	3	$ "31" indicates the number of variables reported once in the non-GF3 area of the station header record.
	6	3	$ "10" indicates the number of variables reported repeatedly in the non-GF3 area of the station header record.
	9	1	$ "I" indicates that data are written in the non-GF3 area of the station header record as integers exclucsively.
	10	8	Blank.
	18	60	Part of FORMAT for reading data in non-GF3 area of the station header record. $ The contents of this field are fixed for STD/CTD users. This field is for the benefit of GF3 users not acquainted with the STD/CTD implementation. See attached form for contents.
	78	3	"001" card sequence number.
2	1	1	"3" record ID.
	2	16	Blank.
	18	60	More of format for reading data in non GF3 area of the station header record. $ Contents are fixed for STD/CTD users. See attached form for contents.
	78	3	"002" card sequence number.
3	1	1	"3" card sequence number.
	2	76	Blank.
	78	3	"003" card sequence number.

A.3.3 STATION HEADER DEFINITION RECORD #1 Continued
RECORD ID 3

Card No.	Card Byte No.	Number of Bytes in Field	Description of Field and Comments
4-24	1	1	"3".
	2	76	$ Contents specified. Identifies header parameter, units, default value and scaling factors. See attached form for contents.
	78	3	Card sequence number "004" to "024".

STATION HEADER DEFINITION RECORD #2

Card No.	Card Byte No.	Number of Bytes in Field	Description of Field and Comments
1	1	1	"3" record ID.
	2	1	Record ID of next record on tape $ it will be a file header record, so set to "5".
	3	75	Blank.
	78	3	"025" card sequence number.
2-3	1	1	"3" record ID.
	2	76	Blank.
	78	3	"026" - "027" card sequence number.
4-22	1	1	"3" record ID.
	2	76	$ Contents specified. More header parameters and classicially measured parameters.
	78	3	"028" - "046" card sequence number.
23-24	1	1	"3" record ID.
	2	76	Blank.
	78	3	"047" - "048" card sequence number.

Figure A-5

GF3 Coding Form — STATION HEADER DEFINITION RECORD #1

RECORD ID	NEXT REC	NO OF HEADER PARAMETERS	NO OF DATA CYCLE PARAMETERS	FORMAT	BLANKS	FORTRAN FORMAT DESCRIPTION (PARTS 2,3 & 4)	CARD SEQUENCE NUMBER
3	3	31	10	I		(1X,I1,9I2,I5,2I4,3I2,3I3,I4,4(I3,I2),I1,2X,I2,3X,80X,	001
3						35(I4,2I5,3I4,4I3),30X)	002
3							003

REC ID	BLANK	PARAMETER CODE	PARAMETER DESCRIM	NAME OF PARAMETER AND UNITS (Plain Language)	MODE	FIELD LENGTH	DUMMY VALUE CODE	SCALE 1 (×)	SCALE 2 (+)	USAGE FLAG	BLANK	SECONDARY PARAMETER CODE	SECOND PARAMETER DESCRIM	CARD SEQUENCE NUMBER
3		101		TRACE CØDE	I		1	1.0	0.0					004
3		102		WATER CØLØR, FØREL-ULE CØDE	I		-91	1.0	0.0					005
3		103		WATER TRANSPARENCY, METERS	I		-91	1.0	0.0					006
3		104		WAVE DIRECTION, WMØ CØDE 885	I		-91	1.0	0.0					007
3		105		WAVE HEIGHT, WMØ CØDE 1555	I		-91	1.0	0.0					008
3		106		SEA STATE, WMØ CØDE 3700	I		-91	1.0	0.0					009
3		107		WIND FØRCE, BEAUFØRT CØDE	I		-91	1.0	0.0					010
3		108		WAVE PERIØD, WMØ CØDE 3155	I		-91	1.0	0.0					011
3		109		WIND DIRECTIØN, WMØ CØDE 877	I		-91	1.0	0.0					012
3		110		WIND SPEED, KNØTS	I		-91	1.0	0.0					013
3		111		BARØMETRIC PRESS, MILLIBARS	I		-91	1.0	0.0					014
3		112		DRY BULB TEMPERATURE, DEG C	I		-93	10.0	0.0					015
3		113		WET BULB TEMPERATURE, DEG C	I		-93	10.0	0.0					016
3		114		WEATHER, WMØ CØDE 4501/4677	I		-91	1.0	0.0					017
3		115		CLØUD TYPE, WMØ CØDE 0500	I		-91	1.0	0.0					018
3		116		CLØUD AMØUNT, WMØ CØDE 2700	I		-91	1.0	0.0					019
3		117	1	DRØP RATE(HI GRAD), M/SEC	I		-92	100.0	0.0					020
3		117	2	DRØP RATE(LØ GRAD), M/SEC	I		-92	100.0	0.0					021
3		118		SHIP RØLL PERIØD, SEC	I		-92	10.0	0.0					022
3		119		SALINITY ERR(FRØM T,S), PPT	I		94	1000.0	0.0	A		4		023
3		120	1	VERT. RESØL. T 3 SIG DIG, DBAR	I		93	1.0	0.0	A		2		024

CODED BY DATE PAGE OF

Figure A-6

GF3 Coding Form C.

STATION HEADER **DEFINITION RECORD** #2

Record ID	Next Rec	No of Header Parameters	No of Data Cycle Parameters	Format	Blanks	Fortran Format Description (Parts 2, 3 & 4)	Card Sequence Number
3	5						025
3							026
3							027

Rec ID	Blank	Parameter Code	Parameter Descrim	Name of Parameter and Units (Plain Language)	Mode	Field Length	Dummy Value Code	Scale 1 (×)	Scale 2 (+)	Usage Flag	Blank	Secondary Parameter Code	Second Parameter Descrim	Card Sequence Number
3		120	2	VERT. RESØL. T, PØWER ØF 10	I		92	1.0	0.0	A		2		028
3		121	1	RMS NØISE T 3 SIG DIG, DEG C	I		93	1.0	0.0	A		2		029
3		121	2	RMS NØISE T, PØWER ØF 10	I		92	1.0	0.0	A		2		030
3		120	1	VERT. RESØL. S 3 SIG DIG, DBAR	I		93	1.0	0.0	A		4		031
3		120	2	VERT. RESØL. S, PØWER ØF 10	I		92	1.0	0.0	A		4		032
3		121	1	RMS NØISE S 3 SIG DIG, PPT	I		93	1.0	0.0	A		4		033
3		121	2	RMS NØISE S, PØWER ØF 10	I		92	1.0	0.0	A		4		034
3		122		VERT. RESØL./NØISE CØDE	I		1	1.0	0.0					035
3		123		NUMBER ØF WATER BØTTLES	I		1	1.0	0.0					036
3		1		PRESSURE, DBAR	I		-91	1.0	0.0					037
3		2		TEMPERATURE, DEG C	I		-94	1000.0	0.0					038
3		4		SALINITY, PPT	I		-94	1000.0	0.0					039
3		7		DISSØLVED ØXYGEN, ML/L	I		-93	100.0	0.0					040
3		20		SILICATE, MICRØGRAMATØMS/L	I		-93	10.0	0.0					041
3		21		INØRGANIC PHØSPHATE, DITTØ	I		-93	100.0	0.0					042
3		22		TØTAL PHØSPHØRUS, DITTØ	I		-92	100.0	0.0					043
3		23		NITRITES, DITTØ	I		-92	100.0	0.0					044
3		24		NITRATES, DITTØ	I		-92	10.0	0.0					045
3		16		PH	I		-92	100.0	0.0					046
3														047
3														048

CODED BY | DATE | PAGE OF

A.3.4 FILE (CRUISE) HEADER RECORD
RECORD ID 5

Card No.	Card Byte No.	Number of Bytes in Field	Description of Field and Comments
1	1	1	"5" record identifier.
	2	1	Record identifier of the next record ($ set to "4").
	3	9	Project name.
	12	2	IOC Country Code for the country containing the institution collecting the data in this file (i.e. the original source of data), Table 1.
	14	1	Enter "9" to indicate that following code is National.
	15	3	National Institution Code (if available) of the institute collecting the data in this file (i.e., original source of data).
	18	18	Country - plain language) original source of data as
	36	18	Institute - plain language) coded in bytes 12 - 17.
	54	6	Date (YYMMDD) this version of the file was created (YY = Year, MM = Month, DD = Day).
	60	6	Time (HHMMSS) this version of the file was created (HH = Hours, MM = Minutes, SS = Seconds).
	66	12	Processing Number or Identifier assigned to this file by data center.
	78	3	Card Sequence Number "001".
2	1	1	"5" record identifier.
	2	2	Platform type code (primary platform), Table 3.
	4	8	Platform type - plain language (e.g., Ship, Buoy, Aircraft, Float, etc.).
	12	1	Meaning of following bytes, 1 = ITU, 2 = WMO/IOC, 9 = National or local identifier.
	13	9	Specific Platform Code - e.g., ship or aircraft call sign, mooring or buoy identifier, etc., Table 4.

A.3.4 FILE (CRUISE) HEADER RECORD Continued
RECORD ID 5

Card No.	Card Byte No.	Number of Bytes in Field	Description of Field and Comments
	22	22	Platform name - plain language - commonly used name e.g., ship name.
	44	10	Reference No./Identifier assigned by the Institution originating the data to the Cruise of the platform, during which the data was collected.
	54	12	Date/Time at start of cruise expressed as CCYYMMDDHHMM.
	66	12	Date/Time at end of cruise expressed as CCYYMMDDHHMM CC = Century, YY = Year, MM = Month, DD = Day, HH = Hours, MM = Minutes. Enter to appropriate precision leaving remaining digits blank.
	78	3	Card Sequence Number "002".
3	1		Secondary platform information in same format as card 2, for example a mooring may be the primary platform, in which case it is relevant to specify the ship setting that mooring.
	78	3	Card Sequence Number "003".
4	1	1	"5" record identifier.
	2	14	Date/time of the earliest observation in the file. Expressed as CCYYMMDDHHMMSS where CC = Century, YY = Year, MM = Month, DD = Day, HH = Hours, MM = Minutes and SS = Seconds. Expressed in GMT and entered to the appropriate precision leaving remaining digits blank.
	16	14	Date/time of the latest observation in the file. Expressed as for preceeding byte.
	30	6,1	Latitude expressed as DDMMHH, (N/S) entered only if all observations in this file are collected at the same position - otherwise fill with 9's.
	37	7,1	Longitude expressed as DDDMMHH, (E/W) entered only if all observations in this file are collected at the same position - otherwise fill with 0's. Note: DDor DDD = Degrees, MM = Minutes and HH = Hundreths of a minute.

Card No.	Card Byte No.	Number of Bytes in Field	Description of Field and Comments
	45	3	Positional uncertainty or range of observations in the file about the position entered in bytes 30 - 44 expressed in tenths of a nautical mile.
	48	6	Depth to bottom in tenths of meter below sea level at the position entered in bytes 30 - 44.
	54	6	Depth of observations below sea level in tenths of a meter (height above sea level is expressed as a negative value) - entered only if all observations in the file are collected at the same depth relative to sea level - otherwise filled with 9's.
	60	6	Depth of observation below sea floor in tenths of a meter (height above sea floor is expressed as a negative value) - entered only if all observations in the file are collected at the same depth relative to the sea floor - otherwise filled with 9's. An entry in this field should be accompanied by an entry in bytes 48 - 53 wherever possible.
	66	6	Minimum observation depth of this file in tenths of a meter below sea level (height above sea level expressed as negative value) should be filled with 9's if not known or an entry has been made in bytes 54 - 65.
	72	6	Maximum observation depth of this file in tenths of a meter below sea level (height above sea level expressed as negative value) - should be filled with 9's if not known or an entry has been made in bytes 54 - 65.
	78	3	Card sequence number "004".
5	1	1	"5" record identifier.
	2	1	Flag to define the usage of fields in bytes 2 - 32. Set to "1" if fields are used to define the start and end positions of the file. Set to "2" if fields are used to define the geographic limits within which all observations in the file were collected. Set to "9" if the fields are not used - note that fields not used in bytes 2 - 32 should be filled with 9's.

A.3.4 FILE (CRUISE) HEADER RECORD Continued
RECORD ID 5

Card No.	Card Byte No.	Number of Bytes in Field	Description of Field and Comments
	3	6,1	Start/Southern Latitude DDMMHH, (N/S).
	10	7,1	Start/Western Longitude DDDMMHH, (E/W).
	18	6,1	End/Northern Latitude DDMMHH, (N/S).
	25	7,1	End/Eastern Longitude DDDMMHH, (E/W).
			Note: DD or DDD = Degrees, MM = Minutes and HH = Hundredths of a minute).
	33	5	Code for type of observation in this file Table 5.
	38	1	Validation flag for data in this file Table 6.
	39	12	Identifier assigned to this file by the institute collecting (originating) the data.
	51	6	No. of Series (stations) within this file - fill with 9's if not known.
	57	10	No. of data cycles within this file - fill with 9's if not known.
	67	11	Unassigned - leave blank.
	78	3	Card Sequence number "005".
6-24			Blank.

Figure A-7

GF3 Coding Form D. FILE HEADER RECORD (1st 400 chars.)

ORIGINAL SOURCE OF DATA

RECORD ID	NEXT RECORD	PROJECT NAME	COUNTRY CODE	CODE FLAG	INSTITUTION CODE	NAME OF COUNTRY — ORIGINAL SOURCE OF DATA (Plain Language)	NAME OF INSTITUTION — ORIGINAL SOURCE OF DATA (Plain Language)	DATE FILE WAS CREATED Y Y M M D D	TIME FILE WAS CREATED H H M M S S	PROCESSING NUMBER ASSIGNED TO FILE BY DATA CENTRE	CARD SEQUENCE NUMBER
5	4										001

PRIMARY PLATFORM

RECORD ID	PLATFORM TYPE CODE	PLATFORM TYPE NAME (Plain Language)	CODE FLAG	SPECIFIC PLATFORM CODE	PLATFORM NAME (Plain Language)	ORIGINATOR'S CRUISE IDENTIFIER	DURATION OF CRUISE START DATE/TIME C C Y Y M M D D H H M M	DURATION OF CRUISE END DATE/TIME C C Y Y M M D D H H M M	CARD SEQUENCE NUMBER
5									002

SECONDARY PLATFORM

RECORD ID									CARD SEQUENCE NUMBER
5									003

SPACE/TIME COORDINATES

RECORD ID	START DATE/TIME C C Y Y M M D D H H M M S S	END DATE/TIME C C Y Y M M D D H H M M S S	POSITION (IF FIXED) LATITUDE D D M M H H N/S	POSITION (IF FIXED) LONGITUDE D D D M M H H E/W	POSITIONAL ERROR/RANGE N N T	SEA FLOOR DEPTH M M M M M T	OBSERVATION DEPTH RELATIVE TO SEA LEVEL M M M M M T	OBSERVATION DEPTH RELATIVE TO SEA FLOOR M M M M M T	MINIMUM OBSERVATION DEPTH BELOW SEA LEVEL M M M M M T	MAXIMUM OBSERVATION DEPTH BELOW SEA LEVEL M M M M M T	CARD SEQUENCE NUMBER
5											004

POSITION LIMITS; IDENTIFIERS & COUNTS

RECORD ID	USAGE FLAG	START/SOUTHERN LATITUDE D D M M H H N/S	START/WESTERN LONGITUDE D D D M M H H E/W	END/NORTHERN LATITUDE D D M M H H N/S	END/EASTERN LONGITUDE D D D M M H H E/W	CODE FOR TYPE OF OBSERVATIONS	VALIDATION	IDENTIFIER ASSIGNED TO FILE BY DATA ORIGINATOR	NO OF stations IN FILE	NO OF DATA CYCLES IN FILE	BLANKS	CARD SEQUENCE NUMBER
5												005

Remainder of record is blank.

CODED BY	DATE	PAGE OF

A.3.5 DATA CYCLE DEFINITION RECORD
RECORD ID 4

Card No.	Card Byte No.	Number of Bytes in Field	Description of Field and Comments
1	1	1	"4" Record ID.
	2	1	"6" next record is a Series Header.
	3	3	Header Record indicate the number of parameters that are reported in the data cycle record once per record (typically = 000).
	6	3	Indicate the number of parameters in each data cycle (e.g, for a data cycle consisting of pressure, temperature, conductivity and salinity this field = "004").
	10	8	Blank.
	18	60	*First part of FORTRAN format for reading the last 1900 bytes of the data cycle record.
	78	3	"001" card sequence number.
2	1	1	"4" Record ID.
	2	16	Blank.
	18	60	*Continuation of FORTRAN format for reading the last 1900 bytes of the data cycle record.
	78	3	"002" card sequence number.
3	1	1	"4" record ID.
	2	16	Blanks.
	18	60	*Final part of FORTRAN format for reading the last 1900 bytes of the data cycle record.
	78	3	"003" card sequence number.

* For more instruction on the use of these fields, see Section 4.

A.3.5 DATA CYCLE DEFINITION RECORD Continued
RECORD ID 4

Card No.	Card Byte No.	Number of Bytes in Field	Description of Field and Comments
4	1	1	"4" record identifier.
	2	1	Leave blank - reserved for future use.
	3-10	8	Parameter Code - use standard parameter code as far as possible, Table 7.
	11-13	3	Parameter Discriminator - for use in discriminating between parameters repeated in the same data cycle or in the same "header" area.
	14-40	27	Name of the Parameter and its units - plain language.
	41	1	Mode - Place an "I" in this byte to indicate the parameter is written as an integer number (I = integer).
	42-45	4	Leave blank.
	46-48	3	*Dummy value code. (Code for default values).
	49-56	8	Scale 1(*) Scale 2(*) Scales used to reduce the number of bytes needed to record the parameter. To recover the parameter in the units given in bytes 14 - 40, multiply the recorded parameter value by Scale 1 and then add Scale 2. Set Scale 1 to 1.0 if not used and set Scale 2 to 0.0. Use the following equation to recover the parameter. Real value = value on tape x Scale 1 + Scale 2 where "real value" means the physical value in the units given in bytes 14 - 40.
	65	1	Attribute Flag - set to 'A' if the parameter is used to define the attribute of another parameter - otherwise leave blank.
	66	1	Leave blank - reserved for future use.

* For more instruction on the use of these fields, see Section 4.

A.3.5 DATA CYCLE DEFINITION RECORD Continued
RECORD ID 4

Card No.	Card Byte No.	Number of Bytes in Field	Description of Field and Comments
	67-74	8	Secondary Partameter Code - if the Attribute Flag = 'A' then enter the Parameter Code of the parameter whose attribute is being defined. If the Attribute Flag = (A' and the parameter is a function of 2 other parameters, e.g., cross spectrum then enter the Parameter Code of the 2nd parameter - otherwise leave blank. Use standard parameter codes as far as possible, Table 7.
	75-77	3	Secondary Parameter Discriminator - if the Attribute Flag = 'A' and the parameter whose attribute is being defined itself has a Parameter Discriminator then enter that Parameter Discriminator - otherwise leave blank.
	78-80	3	Card sequence number "004".
	PARAMETER 2		
5	1-80	80	Card 5 same as card 4 but for parameter 2, etc.
	PARAMETER 21		
24	1-80	80	Card 24 (end of record).

PARAMETERS 22-42, 43-63 etc.

Further parameters may be included on succeeding Definition Records if necessary in the same format as above using Cards 28-48 for PARAMETERS 22-42 on the 2nd record, Cards 52-72 for PARAMETERS 43-63 on the 3rd record etc. The 1st three cards of each record used in this way should be formatted as follows:

Cards 25, 49 etc.: byte 1 should contain the same entry as byte 1 of card no. "1", byte 2 should contain the record identifier of the next record, bytes 3-77 should be left blank and bytes 78-80 should contain the card sequence no. ("025", "049", etc.) Cards 26, 27, 50, 51 etc.: as above but byte 2 should be left blank.

NOTE: The total number of Definition Records required is deduced from the sum of the number of "header parameters" and the number of "data cycle parameters" (bytes 3-8 of card 1). Parameters must be entered in the sequence in which they appear in the "user formatted area" being defined. The "header parameters" must appear before the "data cycle parameters".

Figure A-8

GF3 Coding Form C. DEFINITION RECORD

RECORD ID (1)	NEXT REC (2)	NO OF HEADER PARAMETERS (3–5)	NO OF DATA CYCLE PARAMETERS (6–8)	FORMAT (9)	BLANKS (10–17)	FORTRAN FORMAT DESCRIPTION (PARTS 2, 3 & 4) (18–77)	CARD SEQUENCE NUMBER (78–80)
4	6			1			001
4							002
4							003

REC ID (1)	BLANK (2)	PARAMETER CODE (3–10)	PARAMETER DESCRIM (11–13)	NAME OF PARAMETER AND UNITS (Plain Language) (14–40)	MODE (41)	FIELD LENGTH (42–45)	DUMMY VALUE CODE (46–48)	SCALE 1 (x) (49–56)	SCALE 2 (+) (57–64)	USAGE FLAG (65)	BLANK (66)	SECONDARY PARAMETER CODE (67–74)	SECOND PARAMETER DESCRIM (75–77)	CARD SEQUENCE NUMBER (78–80)
4														004
4														005
4														006
4														007
4														008
4														009
4														010
4														011
4														012
4														013
4														014
4														015
4														016
4														017
4														018
4														019
4														020
4														021
4														022
4														023
4														024

CODED BY | DATE | PAGE OF

A.3.6 STATION HEADER RECORD
RECORD ID 6

Card No.	Card Byte No.	Number of Bytes in Field	Description of Field and Comments
1	1	1	"6" record identifier.
	2	1	Record identifier of the next record. $ It is a plain language record, so set to "0".
	3	9	Project name.
	12	2	IOC Country Code for the country containing the institution collecting the data in this series (i.e., the original source of data), Table 1.
	14	1	Enter "9" to indicate that following code is National.
	15	3	National Institution Code (if available) of the institute collecting the data in this series (i.e., original source of data).
	18	18	Country - plain language) original source of data as
	36	18	Institute - plain language) coded in bytes 12 - 17.
	54	6	Date (YYMMDD) this version of the series was created (YY = year, MM = Month, DD = Day).
	60	6	Time (HHMMSS) this version of the series was created (HH = Hours, MM = Minutes, SS = Seconds).
	66	12	Processing Number or Identifier assigned to this series by data center.
	78	3	Card Sequence Number "001".
2	1	1	"6" record identifier.
	2-3	2	Platform type code (primary platform), Table 3.
	4-1	8	Platform type - plain language (e.g., Ship, Buoy, Aircraft, Float, etc.).
	12	1	Meaning of record bytes 13 - 21, 1-ITU, 2-WMO/IOC, 9-National or local identifier.

Card No.	Card Byte No.	Number of Bytes in Field	Description of Field and Comments
	13	9	Specific Platform Code - e.g., ship or aircraft callsign, mooring or buoy identifier, etc., Table 4.
	22	22	Platform name - plain language - commonly used name e.g., ship name.
	44	10	Reference No./Identifier assigned by the Institution originating the data to the Station of the platform, during which the data was collected.
	54	12	Date/Time at start of station expressed as CCYYMMDDHHMM.
	66	12	Date/Time at end of station expressed as CCYYMMDDHHMM.
			CC = Century, YY = Year, MM = Month, DD = Day, HH = Hours, MM = Minutes. Enter to appropriate precision leaving remaining digits blank.
	78	3	Card Sequence Number "002".
3			Secondary platform information in same format as record bytes 81 - 157, for example a mooring may be the primary platform, in which case it is relevant to specify the ship setting that mooring.
	78	3	Card Sequence Number "003".
4	1	1	"6" record identifier.
	2	14	Date/time of start of the series. Expressed as CCYYMMDDHHMMSS where CC = Century, YY = Year, MM = Month, DD = Day, HH = Hours, MM = Minutes and SS = Seconds. Expressed in GMT and entered to the appropriate precision leaving remaining digits blank.
	16	14	Date/time of end of the series. Expressed as for bytes 2 - 15.

A.3.6 STATION HEADER RECORD Continued
RECORD ID 6

Card No.	Card Byte No.	Number of Bytes in Field	Description of Field and Comments
	30	6,1	Latitude expressed as DDMMHH, (N/S) entered only if all observations in this series are collected at the same position - otherwise fill with 9's.
	37	7,1	Longitude expressed as DDDMMHH, (E/W) entered only if all observations in this series are collected at the same position - otherwise fill with 9's. Note: DDor DDD = Degrees, MM = Minutes and HH = Hundreths of a minute.
	45	3	Positional uncertainty or range of observations in the series about the position entered in bytes 30 - 44 expressed in tenths of a nautical mile.
	48	6	Depth to bottom in tenths of meter below sea level at the position entered in bytes 30 - 44.
	54	6	Depth of observations below sea level in tenths of a meter (height above sea level is expressed as a negative value) - entered only if all observations in the series are collected at the same depth relative to sea level - otherwise filled with 9's.
	60	6	Depth of observation below sea floor in tenths of a meter (height above sea floor is expressed as a negative value) - entered only if all observations in the series are collected at the same depth relative to the sea floor - otherwise filled with 9's. An entry in this field should be accompanied by an entry in bytes 48 - 53 wherever possible.
	66	6	Minimum observation depth of this station in tenths of a meter below sea level (height above sea level expressed as negative value) - should be filled with 9's if not known or an entry has been made in bytes 54 - 65.

Card No.	Card Byte No.	Number of Bytes in Field	Description of Field and Comments
	72	6	Maximum observation depth of this station in tenths of a meter below seas level (height above sea level expressed as negative value) - should be filled with 9's if not known or an entry has been made in bytes 54 - 65.
	78	3	Card sequence number "004".
5	1	1	"6" record identifier.
	2	1	Flag to define the usage of fields in bytes 3 - 32. Set to "1" if fields are used to define the start and end positions of the series. Set to "2" if fields are used to define the geographic limits within which all observations in the series were collected. Set to "9" if the fields are not used - note that fields not used in bytes 3 - 32 should be filled with 9's.
	3	6,1	Start/Southern Latitude DDMMHH, (N/S).
	10	7,1	Start/Western Longitude DDDMMHH, (E/W).
	18	6,1	End/Northern Latitude DDMMHH, (N/S).
	25	7,1	End/Eastern Longitude DDDMMHH, (E/W)
			Note: DD or DDD = Degrees, MM = Minutes and HH = Hundredths of a minute).
	33	5	Code for type of observation in this series Table 5.
	38	1	Validation flag for data in this series Table 6.
	39	12	Identifier assigned to this station by the institute collecting (originating) the data.
	51	6	Fill with 9's.

A.3.6 STATION HEADER RECORD Continued
Record ID 6

Card No.	Card Byte No.	Number of Bytes in Field	Description of Field and Comments
	57	10	No. of water bottles at this station whose data are reported in this station Header Record.
	67	11	Unassigned - leave blank.
	78	3	Card Sequence number "005".
$ 6	1	1	RECORD ID = 6.
	2	1	TRACE (0 = N/A, 1 = Down, 2 = Up, 3 = Average of up and down). Default = 0.
	3	2	WATER COLOR, Forel-ule Code. Default = -9.
	5	2	WATER TRANSPARENCY, meters Secchi disc. Default = -9.
	7	2	WAVE DIRECTION, WMO Code 0885. Default = -9.
	9	2	WAVE HEIGHT, WMO Code 1555. Default = -9.
	11	2	SEA STATE, WMO Code 3700. Default = -9.
	13	2	WIND FORCE, Beaufort scale. Default = -9.
	15	2	WAVE PERIOD, seconds. Default = -9.
	17	2	WIND DIRECTION, degrees east from north. Default = -9.
	19	2	WIND SPEED, knots. Default = -9.
	21	5	BAROMETRIC PRESSURE, millibars. Default = -9.
	26	4	DRY BULB TEMPERATURE, degrees Celsius. Default = -999.
	30	4	WET BULB TEMPERATURE, degrees Celsius. Default = -999.

A.3.6 STATION HEADER RECORD Continued

RECORD ID 6

Card No.	Card Byte No.	Number of Bytes in Field	Description of Field and Comments
	34	2	WEATHER, WMO Code 4501 or 4677. Default = -9.
	36	2	CLOUD TYPE, WMO Code 0500. Default = -9.
	38	2	CLOUD AMOUNT, WMO Code 2700. Default = -9.
	40	3	DROP RATE IN HIGH GRADIENT LAYERS, cm per sec. Default = -99.
	43	3	DROP RATE IN LOW GRADIENT LAYERS, cm per sec. Default = -99.
	46	3	SHIP ROLL DOMINANT PERIOD, seconds x 10. Default = -99.
	49	4	SALINITY OFFSET FROM HISTORICAL POT. TEMP., S TEST, ppm (NODC supplied). Default = 9999.
	53 56	3 2	3 significant places (Default = 999) and power of ten (Default = 99) for EFFECTIVE VERTICAL RESOLUTION OF T, decibars (NODC supplied).
	59 61	3 2	3 significant places (Default = 999) and power of ten (Default = 99) for RMS TEMPERATURE NOISE, degrees Celsius (NODC supplied).
	63	3	3 significant places (Default = 999) and power of ten (Default = 99) for EFFECTIVE VERTICAL RESOLUTION of S, decibars (NODC supplied).
	68 71	3 2	3 significant places (Default = 999) and power of ten (Default = 99) for RMS SALINITY NOISE, parts per thousand (NODC supplied).
	73	1	Code = 0 if test not done, = 1 if test performed on this station, = 2 if test performed only on other stations on cruise (NODC supplied).
	74	2	Blank.
	76	2	Water bottle Code: = 0 if no bottles; = 1 if on separate Nansen cast; = 2 if on Nansen cast on conducting cable; = 3 if Rosette downtrace; = 4 if Rossette uptrace; = 5 if Rossette mixed trace.

A.3.6 STATION HEADER RECORD Continued
RECORD ID 6

Card No.	Card Byte No.	Number of Bytes in Field	Description of Field and Comments
	78	3	"006" card sequence number.
$ 7	1	1	"6" Record ID.
	2	76	Blank.
	78	3	"007" Card sequence number.

$8-24 $ The 1360 bytes on these 17 card images are not ordered by card image. Rather, they are divided up into 35 groups of 38 bytes each (with 30 unused bytes at the end of the record). The 38 bytes in each group have the following assignments.

	1	4	PRESSURE at water bottle sample in decibars (or DEPTH in meters). Default = -9.
	5	5	TEMPERATURE at water bottle sample in millidegrees C. Default = -9999.
	10	5	SALINITY of water bottle sample in parts per million. Default = -9999.
	15	4	OXYGEN of water bottle sample in milliliters per liter. Default = -999.
	19	4	SILICATE of water bottle sample in micro gram atoms per liter times 10. Default = -999.
	23	4	INORGANIC PHOSPHATE of water bottle sample in microgram atoms per liter times 100. Default = -999.
	27	3	TOTAL PHOSPHOROUS of water bottle sample in microgram atoms per liter times 100. Default = -99.
	30	3	NITRITE of water bottle sample in microgram atoms per liter times 100. Default = -99.
	33	3	NITRATE of water bottle sample in microgram atoms per liter times 10. Default = -99.
	36	3	pH of water bottle sample times 100. Default = -99.

Figure A-9

GF3 Coding Form D. (STATION) SERIES HEADER RECORD (1st 400 chars.)

ORIGINAL SOURCE OF DATA

RECORD ID	NEXT RECORD	PROJECT NAME	COUNTRY CODE	CODE FLAG	INSTITUTION CODE	NAME OF COUNTRY – ORIGINAL SOURCE OF DATA (Plain Language)	NAME OF INSTITUTION – ORIGINAL SOURCE OF DATA (Plain Language)	DATE SERIES WAS CREATED Y Y M M D D	TIME SERIES WAS CREATED H H M M S S	PROCESSING NUMBER ASSIGNED TO SERIES BY DATA CENTRE	CARD SEQUENCE NUMBER
6	0										001

PRIMARY PLATFORM

RECORD ID	PLATFORM TYPE CODE	PLATFORM TYPE NAME (Plain Language)	CODE FLAG	SPECIFIC PLATFORM CODE	PLATFORM NAME (Plain Language)	ORIGINATOR'S station IDENTIFIER	DURATION OF station START DATE/TIME C C Y Y M M D D H H M M	END DATE/TIME C C Y Y M M D D H H M M	CARD SEQUENCE NUMBER
6									002

SECONDARY PLATFORM

RECORD ID									CARD SEQUENCE NUMBER
6									003

SPACE/TIME COORDINATES

RECORD ID	START DATE/TIME C C Y Y M M D D H H M M S S	END DATE/TIME C C Y Y M M D D H H M M S S	POSITION (IF FIXED) LATITUDE D D M M H H N/S	LONGITUDE D D D M M H H E/W	POSITIONAL ERROR/RANGE N N T	SEA FLOOR DEPTH M M M M M T	OBSERVATION DEPTH RELATIVE TO SEA LEVEL M M M M M T	OBSERVATION DEPTH RELATIVE TO SEA FLOOR M M M M M T	MINIMUM OBSERVATION DEPTH BELOW SEA LEVEL M M M M M T	MAXIMUM OBSERVATION DEPTH BELOW SEA LEVEL M M M M M T	CARD SEQUENCE NUMBER
6											004

POSITION LIMITS; IDENTIFIERS & COUNTS

RECORD ID	USAGE FLAG	START/SOUTHERN LATITUDE D D M M H H N/S	START/WESTERN LONGITUDE D D D M M H H E/W	END/NORTHERN LATITUDE D D M M H H N/S	END/EASTERN LONGITUDE D D D M M H H E/W	CODE FOR TYPE OF OBSERVATIONS	VALIDATION	IDENTIFIER ASSIGNED TO SERIES BY DATA ORIGINATOR		NO OF water bottles IN station	BLANKS	CARD SEQUENCE NUMBER
6									999999			005

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80

CODED BY	DATE	PAGE OF

Figure A-10

Card	Fields (card columns 1–80)
REC ID 6, card sequence 006	REC ID; WATER [illegible]; WATER COLOR; TRANS-PARENCY; WAVE DIRECTION; WAVE HEIGHT; SEA STATE; WIND FORCE; WAVE PERIOD; WIND DIRECTION; WIND SPEED; BAROMETRIC PRESSURE; DRY BULB TEMPERATURE; WET BULB TEMPERATURE; WEATHER; CLOUD TYPE; CLOUD AMOUNT; DROP RATE HI GRAD; DROP RATE LO GRAD; SHIP ROLL PERIOD; SALINITY COMPUTED FROM THEORETICAL T-S CURVES; VERTICAL TEMPERATURE RESOLUTION (3 SIG-NIFICANT DIGITS, EXPONENT BASE 10); RMS TEMPERATURE NOISE (3 SIG-NIFICANT DIGITS, EXPONENT BASE 10); VERTICAL SALINITY RESOLUTION (3 SIG-NIFICANT DIGITS, EXPONENT BASE 10); RMS SALINITY NOISE (3 SIG-NIFICANT DIGITS, EXPONENT BASE 10); [illegible]; [illegible]; CARD SEQUENCE NUMBER
REC ID 6, card sequence 007	PRES-SURE (DBAR); TEMPER-ATURE (MILLI-DEG C); SALIN-ITY (PPM); OXYGEN (ml/l x 100); SILICATE (µgm·at/LITER); INORG. PHOS-PHATE (x 100); TOTAL PHOS. (x 100); NITRATES (x 100); NITRITES (x 10); PH (x 100); CARD SEQUENCE NUMBER

Repeat this sequence for the number of bottles (up to 35)

CODED BY | DATE | PAGE OF

A.3.7 $ PLAIN LANGUAGE RECORD(S)
THAT FOLLOWS STATION HEADER RECORD

Card No.	Card Byte No.	Number of Bytes in Field	Description of Field and Comments
1	1	1	"0" (zero). Identifies this as a Plain Language Record.
	2	1	Next record ID. Identifies the type of GF3 record to follow.
	3	31	"DOCUMENTATION FOR STATION DATA".
	34	44	Blanks.
	78	3	"001" Card sequence marks.
2	1	1	"0" (zero).
	2	12	Platform name.
	14	12	Cruise name.
	26	08	Station name.
	34	44	Purpose of cruise.
	78	3	"002".
3	1	1	"0" (zero).
	2	38	Institution name.
	40	23	Investigator name. Name of person responsible for the data set.
	64	3	"TEL".
	67	11	Telephone number to reach the investigator named.
	78	3	"003".
4	1	1	"0" (zero).
	2	76	Mailing address to reach the investigator named.
	78	3	"004".

A.3.7 $ PLAIN LANGUAGE RECORD(S) Continued
THAT FOLLOWS STATION HEADER RECORD

Card No.	Card Byte No.	Number of Bytes in Field	Description of Field and Comments
5	1	1	"0" (zero).
	2	16	"UNDERWATER UNIT".
	18	25	Manufacturer name.
	43	12	Model #.
	55	23	Write comments on deployment (e.g., coupled to ship motion, free fall, yo-yo).
	78	3	"005".
6	1	1	"0" (zero).
	2	10	"DIGITIZER:"
	12	20	Digitizer manufacturer name.
	32	8	Model #.
	40	10	"RECORDER:"
	50	20	Recorder manufacturer name.
	70	8	Model #.
	78	3	"006".
7	1	1	"0" (zero).
(19,31*,43*,etc)	2	25	Parameter name and units.
parameter #1 on Card 7	27	15	"TARGET ACCURACY".
parameter #2 on Card 19	42	10	The accuracy which the original investigator feels he achieved in the reported values of of this parameter in the units given above.
parameter #3 on Card 31*			
parameter #4 on Card 43*	52	16	"TARGET PRECISION".
etc.			

* NOTE: Card number greater than 24 must be continued on following plain Language Records.

Card No.	Card Byte No.	Number of Bytes in Field	Description of Field and Comments
	68	10	The precision which the original investigator feels he achieved in the reported values of this parameter in the units given above.
	78	3	"007" or "019" or etc.
8	1	1	"0" (zero).
(20,32,44,etc)	2	22	"SENSOR SERIAL NUMBERS:"
	24	26	Serial numbers of sensor(s) used at this station for this parameter, separated by commas.
	50	23	"PERCENT RAW DATA LOSS:"
	73	5	Estimate of the data which could be measured but were not included in the recorded values, in percent.
	78	3	"008" or "020" or etc.
9	1	1	"0" (zero).
(21,33,45,etc)	2	29	" RAW DATA. SAMPLING INTERVAL:"
	31	10	Time period over which raw sample is determined, in seconds.
	41	20	" SEC. LOGGING RATE:"
	61	10	Rate at which samples are recorded in samples per second.
	71	7	"SEC** - 1".
	78	3	"009" or "021" or etc.

Card No.	Card Byte No.	Number of Bytes in Field	Description of Field and Comments
10	1	1	"0" (zero)
(22,34,46, etc)	2	21	"LAG CORRECTION. STEP"
	23	3	Indicate the sequence of this step during the processing procedure by a natural number increasing chronologically. If no lag correction is performed on this parameter write. " NO".
	26	6	" REF: "
	32	30	Author, date for reference of algorithm used. Full reference must appear in Plain Language Records following the Tape Header Record.
	62	3	"AV:"
	65	5	The time over which data were smoothed to find time derivative for lag correction (if appropriate to the algorithm used) in milliseconds
	70	4	"TAU:"
	74	4	The sensor time lag used in the lag correction (if appropriate to the algorithm used) in milliseconds.
	78	3	"010" or "022" or etc.
11	1	1	"0" (zero)
(23,34,47, etc)	2	17	"DERIVATION. STEP "
	19	3	Indicate the sequence of this step during the processing procedure by a natural number increasing chronologically. If the parameter is not derived from other measured parameters write " NO".

A.3.7 $ PLAIN LANGUAGE RECORD(S) Continued
THAT FOLLOWS STATION HEADER RECORD

Card No.	Card Byte No.	Number of Bytes in Field	Description of Field and Comments
	22	5	" REF:"
	27	35	Author, date of reference for algorithm used. Full reference must appear in Plain Language Records following the Tape Header Record.
	62	5	"FROM:"
	67	11	List abbreviations of all measured parameters from which present parameter is derived, separate by comma's.
	78	3	"011" or "023" or etc.
12	1	1	"0" (zero)
(24,36,48, etc)	2	14	"EDITING. STEP "
	16	3	Indicate the sequence of this step during the processing procedure by a natural number increasing chronologically. If no editing is performed on this parameter write "NO".
	19	5	"REF:"
12 (con't)	24	32	Author, date for reference of algorithm used. Full reference must appear in the Plain Language Records following the Tape Header Record.
	56	16	"PERCENT DELETED:"
	72	5	Give estimate of the amount of data deleted from this station by editing routine, in percent.
	77	1	Blank.
	78	3	"012" or "024" or etc.
13	1	1	"0" (zero).
(25,37,49 etc)	2	16	"SMOOTHING STEP".

AD-A091 866   SCIENCE APPLICATIONS INC MCLEAN VA   F/G 8/10

SPECIFICATIONS FOR AN STD/CTD SYSTEM AT THE NODC,(U)

JUN 80   E J MOLINELLI, R STIEGLITZ   N00014-79-C-0906

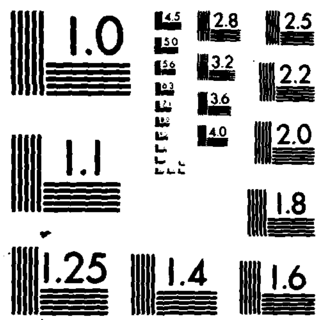
UNCLASSIFIED   SAI-81-195-WA   NL

END

DATE FILMED

81-1

DTIC



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

Card No.	Card Byte No.	Number of Bytes in Field	Description of Field and Comments
	18	3	Indicate the sequence of this step during the processing procedure by a natural number increasing chronologically. If no smoothing (other than that associated with the lag correction) is performed on this parameter write "NO".
	21	5	" REF:"
	26	31	Author, date for reference of algorithm used. Full reference must appear in the Plain Language Records following the Tape Header Record.
	57	6	"SCALE:"
	63	15	Indicate the scale and units over which data were smoothed (E.G. 10 Sec).
	78	3	"013" or "025" or etc.
14	1	1	"0" (zero)
(26,38,50, etc)	2	20	"INTERPOLATION. STEP "
	22	3	Indicate the sequence of this step during the processing procedure by a natural number increasing chronologically. If no interpolating is performed write "NO".
	25	5	" REF:"
	30	24	Author, date for reference of algorithm used. Full reference must appear in the Plain Language Records following the Tape Header Record.
	54	18	"PERCENT INTERPOLATED:"

PLAIN LANGUAGE RECORD(S) Continued
THAT FOLLOW STATION HEADER RECORD

Card No.	Card Byte No.	Number of Bytes in Field	Description of Field and Comments
	72	5	Indicate amount of reported data that are generated by interpolation, in percent.
	77	1	Blank.
	78	3	"014" or "026" or etc.
15	1	1	"0" (zero).
(27,39,51, etc)	2	18	"CALIBRATION. STEP ".
	20	3	Indicate the sequence of this step during the processing procedure by a natural number increasing chronologically. If no calibration is performed for this parameter write "NO".
	23	11	" DATA FROM:"
	34	8	Day, month, year (separated by slashes) on which laboratory calibration was performed.
	42	1	","
	43	34	Institution at which the laboratory calibration was performed.
	77	1	","
	78	3	"015" or "027" or etc.
16	1	1	"0" (zero).
(28,40,52 etc)	2	3	Blank.
	5	9	"STANDARD:"
	14	25	Name standard used at laboratory, e.g. IPTS '48, salt bath, etc.

PLAIN LANGUAGE RECORD(S) Continued
THAT FOLLOW STATION HEADER RECORD

Card No.	Card Byte No.	Number of Bytes in Field	Description of Field and Comments
	39	9	"ACCURACY:"
	48	10	Indicate the accuracy as determined at calibration, units of parameter given above, if unknown leave blank.
	58	10	"PRECISION:"
	68	10	Indicate the precision as determined at calibration, units of parameter given above; if unknown leave blank.
	78	3	"016" or "028" or etc.
17	1	1	"0" (zero)
(29,41,53)	2	23	"OTHER COMPARISON. STEP "
	25	3	Indicate the sequence of this step during the processing procedure by a natural number increasing chronologically. If no other comparisons (e.g. to classical measurements or historical θ,S curves) are made write " NO".
	28	24	" NUMBER OF COMPARISONS:"
	52	4	Indicate the number of comparisons made to determine a correction to the sensor that is applied to the data reported for this station.
	56	9	"STANDARD:"
	65	13	Indicate the standard against which comparisons are made, e.g., reversing thermometer, historical potential T,S etc., abbreviate as needed.
	78	3	"017" or "029" or etc.

A.3.7 PLAIN LANGUAGE RECORD(S) Continued
THAT FOLLOW STATION HEADER RECORD

Card No.	Card Byte No.	Number of Bytes in Field	Description of Field and Comments
18	1	1	"0" (zero).
(or 30,42,54, etc)	2	3	Blank.
	5	25	Continue indicating standard
	30	14	"AV. DIFFERENCE:"
	44	10	Indicate the mean difference between the sensor value and the standard value, in units of parameter given above.
	54	14	"RMS DEVIATION:"
	68	10	Indicate the rms deviation of the differences between the sensor value and the standard value, in units of parameter given above.
	78	3	"018" or "030" or etc.

Figure A-11

GF3 Coding Form **PLAIN LANGUAGE RECORD** FOLLOWING STATION HEADER RECORD

RECORD ID	NEXT REC	Description of collection and processing procedures (1st 6 cards in 1st record)	CARD SEQUENCE NUMBER
0		DØCUMENTATIØN FØR STATIØN DATA	001
0			002
0		, TEL	003
0			004
0	U	NDERWATER UNIT:	005
0	D	IGITIZER: RECØRDER:	006
0			*
0			*
0			*
0			*
0			*
0			*
0			*
0			*
0			*
0			*
0			*
0			*
0			
0			
0			
0			
0			
0			

* parameter sequence (see next form).

CODED BY | DATE | PAGE OF

Figure A-12

GF3 Coding Form

PLAIN LANGUAGE RECORD(S) FOLLOWING STATION HEADER RECORD.

RECORD ID (1)	Description of collection and processing procedures for each parameter (2–77)	CARD SEQUENCE NUMBER (78–80)
0	TARGET ACCURACY TARGET PRECISION	
0	SENSOR SERIAL NUMBERS: PERCENT RAW DATA LOSS:	
0	RAW DATA. SAMPLING INTERVAL: SEC. RECORDING RATE: SEC**-1	
0	LAG CORRECTION. STEP REF: AV: TAU:	
0	DERIVATION. STEP REF: FROM:	
0	EDITING. STEP REF: PERCENT DELETED:	
0	SMOOTHING. STEP REF: SCALE:	
0	INTERPOLATION. STEP REF: PERCENT INTRPLTED:	
0	CALIBRATION. STEP DATA FROM:	
0	STANDARD: ACCURACY: PRECISION:	
0	OTHER COMPARISON. STEP NUMBER OF COMPARISONS: STANDARD:	
0	AV. DIFFERENCE: RMS DEVIATION:	
0	TARGET ACCURACY TARGET PRECISION	
0	SENSOR SERIAL NUMBERS: PERCENT RAW DATA LOSS:	
0	RAW DATA. SAMPLING INTERVAL: SEC. RECORDING RATE: SEC**-1	
0	LAG CORRECTION. STEP REF: AV: TAU:	
0	DERIVATION. STEP REF: FROM:	
0	EDITING. STEP REF: PERCENT DELETED:	
0	SMOOTHING. STEP REF: SCALE:	
0	INTERPOLATION. STEP REF: PERCENT INTRPLTED:	
0	CALIBRATION. STEP DATA FROM:	
0	STANDARD: ACCURACY: PRECISION:	
0	OTHER COMPARISON. STEP NUMBER OF COMPARISONS: STANDARD:	
0	AV. DIFFERENCE: RMS DEVIATION:	

CODED BY	DATE	PAGE OF

* This sequence is to be repeated for every data parameter including water bottle parameters. Begin with card seq. no. 7 and use as many as are needed.

A.3.8 DATA CYCLE RECORD
RECORD ID 7

Card No.	Card Byte No.	Number of Bytes in Field	Description of Field and Comments
1	1	1	"7" Record ID.
	2	1	Record ID of next record, typically another Data Cycle Record.
	3	4	The number of data cycles in this record. This number is less than or equal to the maximum number of cycles per record. The maximum # cycles = greatest integer in the ratio of 1900 bytes to the number of bytes per cycle.
	7	9	The number of data cycles in this station preceeding the present record. Should be equal to the product of the number of data cycle records preceeding times the maximum number of cycles per record.
	16	5	The record count. Indicate which data cycle record of the present station, the present record is.

The 1900 bytes remaining in this record are not in card image but contain an integral number of data cycles. Each data cycle begins immediately after the previous cycle ends. For the number of parameters in the cycle and their arrangement within the 1900 byte field one must see the Data Cycle Definition Record generated by the supplier.

GF3 Coding Form E. E1; DATA CYCLE RECORD

Figure A-13

FORM E1: Except for first 20 characters,
the form is user formatted
according to corresponding
Data Cycle Definition Record.

1	2	3–6	7–15	16–20
RECORD ID	NEXT REC	NO OF DATA CYCLES IN RECORD	NO OF PRECEDING DATA CYCLES	DATA CYCLE RECORD COUNT
7				

USER FORMATTED AREA

CODED BY	DATE	PAGE OF

A.3.9 END OF TAPE RECORD ID 8

Card No.	Card Byte No.	Number of Bytes in Field	Description of Field and Comments
1	1	1	"8" Record ID.
	2	1	"1" if data are continued on another tape, "9" otherwise.
	3	10	Set to 9's.
	13	12	Name or number of following tape on which data are continued. If data are not continued set to 9's.
	25	53	Set to 9's.
	78	3	"001" Card sequence number.
2-24	1	1	"8" record ID.
	2	76	Comments in plain language. $ leave blank.
	78	3	"002", "003" etc. Card sequence number.

GF3 Coding Form F.

END OF TAPE RECORD

RECORD ID (1)	NEXT REC (2)	SET TO 9's (3–12)	NAME OR NUMBER OF FOLLOWING TAPE IF FILE IS CONTINUED – OTHERWISE SET TO 9's (13–24)	SET TO 9's (25–77)	CARD SEQUENCE NUMBER (78–80)
8		9999999999		99999999999999999999999999999999999999999999999999999	001

REC ID (1)	PLAIN LANGUAGE COMMENTS (2–77)	CARD SEQ NO (78–80)
8		002
8		003
8		004
8		005
8		006
8		007
8		008
8		009
8		010
8		011
8		012
8		013
8		014
8		015
8		016
8		017
8		018
8		019
8		020
8		021
8		022
8		023
8		024

CODED BY	DATE	PAGE OF

Figure A-14

A.4 GUIDANCE NOTES ON USE OF GF3 $ STD/CTD IMPLEMENTATION

A.4.1 Guidance Note (1): Record Identifier or Next Record and Its Effect on File Independence

A.4.1.1 Record Identifier of Next Record

Certain computer systems require that the format of a record should be known before it is actually read and, in order that this requirement may be met without affecting the versatility of the GF3 format, the second character in each record is reserved for the record identifier of the record that follows it. It is essential that this character is entered correctly. Please note that EOF marks are not considered as records in this context, and that the record preceding an EOF should contain the identifier of the record immediately following the EOF. If the file is continued on another tape then the End of Tape Record should reference the Tape Header Record of the following tape. If the file is not continued then the End of Tape Record will not be followed by another record and should reference the dummy record identifier '9'.

A.4.1.2 File Independence

One method of finding out which type of record ads a specific GF3 Data File is to examine the second character (next record identifier) of the last record of the previous file (whether it be a Tape Header File or another Data File). If this method is used then individual Data

Files cannot be considered independently of each other and this may cause problems with certain operating systems. If such is the case then a more appropriate method is obtained by recognizing that all Data Files on a given GF3 tape must start with the same record type. On a given tape this record type can be identified by examining the second character of the last record of the Tape Header File. Thus, on reading a GF3 tape one can in fact have Data File independence i.e. in reading a given Data File one does not need to access records in neighboring Data Files. On writing a GF3 tape and in particular the second character of the last record in each Data File one only need know in advance whether the file is to be followed by another Data File or A Tape Terminator File - in the former case one sets it as for the second character of last record in the Tape Header File and in the latter to '8' to identify the following End of Tape Record.

It is inherent in the use of GF3 as an exchange format that Data Files are not used independently of the contents of the Tape Header File. In an archive situation this should not introduce problems if the original exchange tapes themselves are archived. However, if one wishes to reorganize the data files (e.g., to reduce the number of tapes) then of course the lack of file independence from the Tape Header File may be somewhat inconvenient, but it should not be too difficult at that stage to merge the Tape Header information into the individual Data Files. The Tape Header Record itself can be reformatted into a Plain Language Record and, together with any Plain Language Records and Definition Records in the Tape Header File, copied into the file heading of each Data File.

A.4.2 Guidance Note (2): USER FORMATTED AREA - DATA CYCLE PARAMETERS

The last 1900 bytes of the data cycle record constitute a "user formatted area". Within any "user formatted area" two types of parameter are recognized i.e. "header parameters" and "data cycle parameters" - the "header parameters" are defined as those parameters that occur only once in the "user formatted area" of each record - the "data cycle parameters" are those that are repeated as many times as there are data cycles in the record i.e. each "data cycle parameter" must be repeated for each data cycle.

Please note - in any "user formatted area" all the "header parameters" if any, must precede the "data cycle parameters"

The strength of GF3 lies in the flexibility with which the user can define "header" and "data cycle" parameters at the Data Cycle Record level. In a relatively simple manner, the format provides a high degree of versatility and enables the user to choose a "solution" that is tailored to his requirements. The simplest solution is provided where one has only one type of parameter - those constant for the series as a whole and those varying with each data cycle in the series. In this case one would only need to define "data cycle parameters" for the "user formatted area" of the Data Cycle Records. However, it is recognized in GF3 that this simple solution is not necessarily the most convenient, least ambiguous or indeed the most efficient way of storing

all types of STD/CTD data, and that there may be occasions where it is more appropriate to include "header parameters" within Data Cycle Records (e.g., time reported once per record while pressure, temperature and conductivity are reported many times per record).

A.4.3 Guidance Note (3): User Formatted Area - FORTRAN Format Description

A.4.3.1 In reading Data Cycle Records with "user formatted area", the recipient of the data needs to know the FORTRAN format of the entire data record (1920 bytes). There are two components of this format:

A: the format for the fixed section of the record i.e., the first 20 bytes in the case of Data Cycle Records - this component is described as Part 1 of the FORTRAN format statement, and can obviously be deduced from the GF3 record specification in Section 3. It is the responsibility of the data recipient to include this format within his reading program - It is not contained within the Definition Records.

B: the format of the "user formatted area" of the record i.e., the last 1900 bytes in the case of Data Cycle Records - the format of this component is contained in Parts 2 - 4 of the FORTRAN format statement, and is included in bytes 18 - 237 of the appropriate Definition Record.

A.4.3.2 FORTRAN Format of the Fixed Section of the Data Cycle Record (i.e., Part 1).

The FORTRAN format description of this section takes the form:

(2I1, I4, I9, I5,

Note - no closing bracket

A.4.3.3 FORTRAN Format of the "user formatted area" of Data Cycle Records (i.e., Parts 2 - 4)

The FORTRAN format description of this section is given in bytes 18 - 237 of the appropriate Definition Record. It must be entered on the Definition Record as though it was only needed for reading in the "user formatted area" i.e., it should start wth an opening bracket "(" and should completely disregard the fixed area of the record being defined. It should obey FORTRAN rules for repetition - all brackets must be paired. 3 x 60 = 180 bytes are allowd in the Definition Record for the format description (i.e., Parts 2 - 4) of the "user formatted area". Probably 60 bytes will usually be sufficient, but the extra space has been allotted to allow for very complex formats e.g., if there is a large number of parameters. The format is terminated by the closing bracket")" that pairs with the opening bracket "(". The remainder of the 180 bytes should be set to blank. If it is necessary to break the format into 60 byte parts, care should be taken not to leave significant blanks at the end of each part. Preferably, terminate each part at a comma",".

The repeat specifications in the format statement must correspond with the number of parameters recorded once per record (Definition Record bytes 3 - 5 - "header parameters") and the number of parameters repeated each data cycle (bytes 6 - 8 - "data cycle parameters").

A.4.3.4 Examples

If in a Data Cycle Record three parameters are recorded once per record (i.e., are "header parameters") with another five in each data cycle (i.e., "data cycle parameters"), the format of the "user formatted area" might be

(3F8.2,50(5F9.1))

or perhaps

```
(2F6.2,F5.3,50(2F6.1,3F11.3))
 2      +1=3    2      +3=5
```

The second format will be used as the basis of the examples that follow. Note that in both cases the repeat count "50" must be large enough to read at least 1900 characters e.g., (2x6+1x5+50x(2x6+3x11)) is greater than 1900 for the second case.

If there were no header parameters, the five parameters repeated in each data cycle could still remain in the Data Cycle Record but would now take on the format

(50(2F6.1,3F11.3))

In no case should individual data cycles be allowed to cross record boundaries i.e., each record should contain an integral no. of data cycles. In reading data cycles from the "user formatted areas" of Data Cycle Records the actual number of data cycles contained in that area is entered as a field in the "fixed area" of the record (bytes 3 - 6 in the Data Cycle Record).

A.4.3.5 FORTRAN Format of the Complete Record

As mentioned in A.4.3.1, in reading Data Cycle Records, the recipient of data needs to know the FORTRAN format of the entire data record (1920 bytes) and not just the format of its separate components. The full format can be constructed by taking the format description for the fixed component from within the program (see A.4.3.2) and attaching to the end of it the *format description* for the "user formatted area" as read off the Definition Record having first removed its opening "(". Thus for the examples in A.4.3.4 the following full FORTRAN format descriptions would be needed for reading the records:

Example 4.3.4

Case 1 Data Cycle Record Format
(2I1,I4,I9,I5,3F8.2,50(5F9.1))

Case 2 Data Cycle Record Format
(2I1,I4,I9,I5,2F6.2,F5.3,50(2F6.1,3F11.3))

A.4.3.6 Recommended Format Types

$ Only FORTRAN format types I, and X are to be used. These are defined as follows:

I: the I format is used for arithmetic values which are whole decimal numbers. It is also used for the names of parameters and variables when these names have been coded as integer numbers. Spaces (blanks) are normally used instead of leading zeros to fill out the specified field. The decimal point is never used.

Examples:

Format	Uses
I5	12345, b1234, bbbb1 ("b" means "blank")
2I4	12341234, bb12222bb12

Trailing blanks should not be used to fill a field because these are generally interpreted as zeros by computers. For example a number written in I4 format as b12b would be read as 0120. I-format numbers should therefore be written right adjusted (written at the right-hand side of the field) with blanks or zeros for fill on the left. An all-blank field will generally be read as 0 (zero) or -0, depending on the computer used.

X: this format is literally unspecified. Fields written with this format will generally consist of blanks but not necessarily so. Characters read under X formats are disregarded.

Note that the scaling factors (Definition Record card bytes 36 - 51) for each parameter may be used to convert non-integers to integers according to the following formula.

$$\text{Value on Tape} = \frac{\text{Real Value - Scale 2}}{\text{Scale 1}}$$

where "real value" means the physical value in the units given in the parameter definition.

Note that the scaling factors themselves are real and can also be used to convert data written on the tape in non standard units to standard units.

Care must be taken that the word lengths of the computers used are sufficient for the resolution needed by each parameter. Again, use of the scaling factors may help.

Whenever parameters with an extremely wide range of values are to be recorded e.g., phyto-plankton data, pollution data etc., one is tempted to use E formats. However, because of the incompatibility of E and D formats on different computers the E and D formats should not be used - the problem may be overcome by recording the mantissa and the exponent of such parameters as separate parameters each written in integer form. (See GF3 parameter code table).

A.4.3.7 $ Automatic Processing of STD/CTD Tapes in GF3

This can be illustrated by considering the processing of a Data Cycle Record in FORTRAN:

a) Define an integer array (IARR) long enough to hold the maximum number of data values there can possibly be in the user formatted area i.e., 1900 words.

b) Read the Data Cycle Definition Record and construct the complete format (see A.4.3.5) in the text array IFORM.

c) Read the data into the integer array with the statement.

```
READ (a,IFORM) ID, NID, NCYC, ICYC, NREC, IARR
```

where a	=	logical unit number of tape drive
ID	=	record identification number of present record
NID	=	record identification number of next record
NCYC	=	number of data cycles in present record
ICYC	=	total number of data cycles preceeding this record
NREC	=	data Cycle Record count for the present record.

The number of useful data words in IARR will be N = NHPAR + (NDPAR*NCYC) where:

NHPAR = number of header parameters in "user formatted area" (contained in bytes 3 - 5 of card 1 of the Data Cycle Definition Record).

NDPAR = number of data cycle parameters in "user formatted area" (contained in bytes 6 - 8 of card 1 of the Data Cycle Definition Record).

A.4.4 Guidance Note (4): Use of the Parameter Discriminator Field

It is expected that most parameters will not require the use of this entry in their definition within Data Cycle Definition Records. However, occasions do arise when the use of the Parameter Code (and the Secondary Parameter Code where appropriate - see 4.6) is by itself insufficient to differentiate between parameters repeated within the same data cycle. The Parameter Discriminator field is used solely for the purposes of discriminating between such parameters.

For example:

A chain of 5 thermistors at 5 different depths recording within the same series may produce data cycles thus:

time, temperature, temperature, temperature, temperature, temperature

In this case each of the five temperature parameters would have associated with it a unique parameter discriminator - e.g., a sequential no. thus 1, 2, 3, 4 and 5 respectively. The 'depth' can then be directly associated with each parameter in turn in the form of a parameter attribute by specifying both the parameter code for temperature and the parameter discriminator (i.e., 1, 2, 3, 4 or 5 respectively) in card bytes 67 - 77 of the Definition Record defining each of the parameter attribute fields.

A.4.5 Guidance Note (5) - Dummy Values

The identification of dummy parameter values which may occur in data sets with different recording frequencies for the parameters (e.g., dissolved oxygen content), or in data sets with missing values for some parameters caused, for example, by instrument malfunction, could create a problem. Most of the standard oceanographic parameters allow the 'all-9's-method' as a unique 'dummy flag' but even this is not without problems in that

i) 'all 9's' may in fact be a reasonable value for a parameter and ii) without examining the FORTRAN format description one does not know how many digits are contained within the stored value e.g. does one test the value against 999 or 9999. To overcome these problems the following method is used to define "dummy values" - each parameter defined in a "user formatted area" has associated with it in the appropriate Data Cycle Definition Record a two digit field "Dummy Value Code" in the range -99 to 99 from which can be derived its "dummy value" as follows:

i) the tens digit defines the dummy digit

ii) the units digit defines the no. of dummy digits in the "dummy value"

iii) the sign of the "Dummy Value Code" defines the sign of the "dummy value".

"Dummy Value Codes" -9, -8...-1,0,2,3,4....9 are meaningless as are Dummy Value Codes with a units digit of 0 e.g., 10, 20, 30 etc., -90, -80 etc. Examples of valid codes are as follows:

"Dummy Value Code"	"dummy value"
1	0
11	1
-11	-1
12	11
-12	-11
13	111
23	222
-23	-222
33	333
32	33
92	99
95	99999
-95	-99999
etc	etc

It is important that the width of the "dummy value" is compatible with the width defined for the parameter field itself in the FORTRAN format description e.g.,

there is no point assigning a "dummy value code" of 95 (99999), -94(-9999) for an I4 field although "dummy value codes" of 94(9999), -93(-999), 93(999) etc. would be acceptable.

Please note The above technique should be applied only for the value of the parameter as stored on the tape i.e., before the factors - Scale 1 and Scale 2 are applied to it. Thus for a Parameter with Scale 1 = 5.0, Scale 2 = 3, "Dummy Value Code" = 93(999) the dummy value stored on tape would be 999 although after conversion it would become 4998.

A.5 STANDARD CODE TABLES

A.5.1 Introduction

The following code tables have been developed for use with the GF3 format as specified in Part I of Appendix 12 to IOC Manuals and Guides No. 9.

Table 1: IOC Country Code

Table 2: GF3 Common Character Set for Magnetic Tape

Table 3: Platform Type Code

Table 4: Specific Platform Code

Table 5: Type of Observations Contained in File or Series

Table 6: Validation Flag

Table 7: Parameter Code (generated for STD/CTD data)

Table 8: Flag Codes (generated for STD/CTD data)

The maintenance and updating of these tables will be provided by the Intergovernmental Oceanographic Commission Working Committee on International Oceanographic Data Exchange on a continuing basis.

TABLE 1: IOC COUNTRY CODE

Code	Country
72	Albania
AL	Algeria
08	Argentina
09	Australia
10	Austria
11	Belgium
13	Bolivia
14	Brazil
15	Bulgaria
12	Burma
18	Canada
19	Ceylon (Sri Lanka)
20	Chile
21	China
22	Colombia
RC	Congo
CR	Costa Rica
CU	Cuba
DA	Dahomey (Benin)
26	Denmark
70	Dominican Republic
28	Ecuador
27	Egypt
75	El Salvador
34	Finland
35	France
96	German Democratic Republic
06	Germany, Federal Republic of
GH	Ghana
36	Greece
37	Guatemala
GU	Guinea
38	Haiti
HO	Honduras
46	Iceland
41	India
86	Thailand
88	Tunisia
89	Turkey
90	Union of Soviet Socialist Republics
74	United Kingdom
42	Indonesia
IN	Intergovernmental/International
45	Ireland
47	Israel
48	Italy
IC	Ivory Coast
JA	Jamaica
49	Japan
24	Korea, Republic of
52	Lebanon
55	Madagascar
MS	Malaysia
57	Mexico
MO	Monaco
56	Morocco
MZ	Mozambique
64	Netherlands
59	New Caledonia
61	New Zealand
NI	Nigeria
58	Norway
62	Pakistan
PA	Panama
65	Peru
66	Philippines
67	Poland
68	Portugal
73	Romania
SE	Senegal
SL	Sierra Leone
91	South Africa
29	Spain
SU	Sudan
77	Sweden
ZA	Tanzania, United Republic of
31	United States of America
99	Unknown
92	Uruguay
93	Venezuela
94	Vietnam, Socialist Republic of
95	Yugoslavia

TABLE 2: GF3 COMMON CHARACTER SET FOR MAGNETIC TAPE

(as entered on Tape Header Record, bytes 162-214)

Byte No.	Character or Symbol	BCD Octal Digits	EBCDIC Hexadecimal Digits	Minsk-32 Binary Bits	ASCII Hexadecimal Digits	See Note
162	1	01	F1	000 0001	31	1
163	2	02	F2	000 0010	32	
164	3	03	F3	000 0011	33	
165	4	04	F4	000 0100	34	
166	5	05	F5	000 0101	35	
167	6	06	F6	000 0110	36	
168	7	07	F7	000 0111	37	
169	8	10	F8	000 1000	38	
170	9	11	F9	000 1001	39	
171	0	12	F0	000 0000	30	
172	=	13	7E	001 0101	3D	2
173	:	15	7A	001 1111	3A	
174	>	16	6E	001 1110	3E	
175	blank(space)	20	40	000 1111	20	3
176	/	21	61	000 1100	2F	
177	S	22	E2	100 1000	53	
178	T	23	E3	011 0010	54	
179	U	24	E4	100 1001	55	
180	V	25	E5	100 1010	56	
181	W	26	E6	100 1011	57	
182	X	27	E7	011 0101	58	
183	Y	30	E8	011 0011	59	
184	Z	31	E9	100 1111	5A	
185	,	33	6B	000 1101	2C	
186	(	34	4D	001 0010	28	4
187	-	40	60	000 1011	2D	
188	J	41	D1	100 0011	4A	
189	K	42	D2	010 1010	4B	
190	L	43	D3	100 0010	4C	
191	M	44	D4	010 1100	4D	
192	N	45	D5	100 0101	4E	
193	O	46	D6	010 1110	4F	
194	P	47	D7	011 0000	50	
195	Q	50	D8	100 0110	51	
196	R	51	D9	100 0111	52	
197	*	54	5C	001 1001	2A	
198	]	55	D0	001 1000	5D	6
199	;	56	5E	001 0110	3B	
200	+	60	4E	000 1010	2B	5
201	A	61	C1	010 0000	41	
202	B	62	C2	010 0010	42	
203	C	63	C3	011 0001	43	
204	D	64	C4	011 1111	44	
205	E	65	C5	010 0101	45	
206	F	66	C6	100 0000	46	
207	G	67	C7	100 0001	47	
208	H	70	C8	010 1101	48	
209	I	71	C9	101 1000	49	7
210	.	73	4B	000 1110	2E	
211	)	74	5D	001 0011	29	
212	[	75	C0	001 0111	5B	6
213	<	76	4C	001 1101	3C	
214	———	77	FF	111 1111	FF	8

NOTES:

1. The characters are in collating sequence for BCD.

2. Some BCD character sets use "#" instead of "=".

3. In punched-card code, a blank is the 2-8 punch for BCD, no punches for EBCDIC, and 0-7-8 punches for Minsk-32.

4. Some BCD character sets use "%" instead of "(".

5. Some BCD character sets use "&" instead of "+".

6. EBCDIC and some ASCII character sets use "{}" instead of "[ ]".

7. The Minsk-32 uses the vertical line, "|", instead of "I".

8. This position specifies the "all 1's" characters to be used in the test file. These characters are not to be used on GF3 data-exchange tapes except in the test file.

Tape Mark (end of file, EOF)*:

7-track tape - a single byte block (record) containing the octal characters "17". The parity is even.

9-track tape, NRZI encoding (EBCDIC and ISO standard) -- a single byte block (record) consisting of the Device Control Character, DC3 ("1" bits in tracks 2, 3, and 8 only.)

*Should EOFs other than these come into use they will be added to this table.

TABLE 3 PLATFORM TYPE CODE

(File Header/Series Header Record, Bytes 82-83, 162-163)

CODE FORM D_1D_2

STATION/ PLATFORM	CODE D1 ↓ D2 →	CHARACTERISTIC 0	1	2	3	4	5	6	7	8	9
UNKNOWN	0	NOT MODIFIED OR OTHERWISE SPECIFIED									
LAND/ SEA FLOOR	1		SEA FLOOR - FIXED	SEA FLOOR - MOBILE	BEACH - INTERTIDAL ZONE	LAND/ ONSHORE - FIXED	LAND/ ONSHORE - MOBILE	OFFSHORE STRUCTURE e.g. OilRig	COASTAL STRUCTURE e.g.Lighthouse pier, rock		
SUBMERSIBLE	2		MANNED - MOBILE	MANNED - DRIFTING	UNMANNED - MOBILE	UNMANNED - TOWED e.g.Batfish					
SHIP	3		RESEARCH SHIP	SHIP OF OPPORTUNITY	SMALL SHORE BASED CRAFT	FIXED POSITION e.g.Lightvessel Weathership					
BUOY/ MOORING	4		SURFACE - MOORED	SURFACE - DRIFTING	SUBSURFACE - MOORED	SUBSURFACE -DRIFTING e.g.Swallow Float	SUBSURFACE -VERTICAL PROFILING				
BALLOON	5		FREE RISING (VERTICAL)	FREE FLOATING (HORIZONTAL)	TETHERED						
AIRCRAFT/ SATELLITE/ ROCKET	6		RESEARCH AIRCRAFT	OTHER AIRCRAFT	ROCKET -NON ORBITING	SATELLITE -GEO- STATIONARY ORBIT	SATELLITE -ORBITING	MANNED SPACECRAFT			
UNASSIGNED	7										
UNASSIGNED	8										
OTHER	9		ICE ISLAND								

TABLE 4: SPECIFIC PLATFORM CODE

For ships with call signs consult the ITU list of ship's call signs.

TABLE 5: TYPE OF OBSERVATIONS CONTAINED IN FILE OR SERIES

Code	Type
11502	STD/CTD
11503	XBT
11504	Nansen-type Cast

TABLE 6: VALIDATION FLAG

(File Header/Series Header Record, byte 358)

0 : no specification

1 : not suspected

2 : suspected

Flag refers to the validation of the file or series - if "2" is specified the plain language records following the file or series header should be consulted for details on the data that is considered suspect.

TABLE 7: $ PARAMETER CODES

Code	Parameter	and suggested units
1	pressure	decibars
2	temperature	degrees Celsius
3	conductivity	millimhos per centimeter
4	salinity	parts per thousand
5	sound speed	meters per second
6	time	seconds
7	dissolved oxygen	milliliters per liter
8	water speed	meters per second
9	water direction	degrees clockwise from North
10	speed north	meters per second
11	speed east	meters per second
12	depth	meters
13	nephelometry	
14	ambient light	
15	electrical field	
16	pH	
17	chloride ion concentration	
18	oxygen sensor current	micro Ampsres
19	oxygen sensor temperature	degrees Celsius
20	silicate	microgram atoms per liter
21	inorganic phosphate	microgram atoms per liter
22	total phosphorous	microgrm atoms per liter
23	nitrites	microgram atoms per liter
24	nitrates	microgram atoms per liter
25	ancillary,faster response, temperature	degrees Celsius
26	platinum thermometer reading	degrees Celsius

TABLE 7: $ PARAMETER CODES Continued

Code	Parameter	and suggested units
101	trace = 1 for down, = 2 for up, = 3 for average of both, 0 = unknown	
102	water color,	Forel-ule Code
103	water transparency,	meters Secchi Disc
104	wave direction,	WMO Code 0885
105	wave height,	WMO Code 1555
106	sea state,	WMO Code 3700
107	wind force,	Beaufort scale
108	wave period,	seconds
109	wind direction,	degrees clockwise from North
110	wind speed,	knots
111	barometric pressure	millibars
112	dry bulb temperature	degrees Celsius
113	wet bulb temperature	degrees Celsius
114	weather	WMO Code 4501 or 4677
115	cloud type	WMO Code 0500
116	cloud amount	WMO Code 2700
117	drop rate,	meters per second
118	ship roll period	seconds
119	error (bias)	
120	vertical resolution	decibars
121	rms noise	
122	vertical resolution and noise code	= 1 if present station included, = 2 if present station not included
123	water bottle count	
201	mantissa	
202	exponent	

TABLE 8: $ SUGGESTED FLAG CODES FOR USE WITH DATA EXCHANGED VIA GF3

Flag Code	Flag Meaning	
100	Instability	Density Max
200	Instability	Density Min
1	Depth/Pressure	
2	Temperature	
3	Conductivity/Salinity	
4	Temperature & Conductivity/Salinity	
5	Other	
6	Other & Temperature	
7	Other & Conductivity/Salinity	
8	Other & Temperature & Conductivity/Salinity	
9	All Parameters	
10	Interpolated	
20	Questionable	(by originator)
30	Questionable	(by NODC)
40	Questionable	(by both)

APPENDIX B

THE CALCULATION OF STORAGE REQUIREMENTS

B.1 INTRODUCTION

Here are discussed some of the details of utilizing storage devices for the data. A method for the number of bits in storage that contain no information (i.e., blanks or redundant digits) is presented. The basic approach is to "pack" information-containing bits in sequence without regard to word boundaries. This avoids the wasteful padding with zeros that is necessary to maintain fixed word sizes.

First it is demonstrated that the cost of unpacking such data is negligible compared to the cost of transfering the data from tape to computer memory. It is understood that such packing saves both space on the tape and the number of I/O operations required to transfer the data.

Next is described the delta format, which reduces to a near minimum the number of bits required to specify an array of parameter values. Then an example is given of the amount of tape space saved if the delta format were applied to all the STD/CTD data estimated to reside within the oceanographic commmunity.

The following section describes the station header information requirements. This must be consistent with the information required in the exchange format. How this information might be stored on the permanent tape is discussed here.

Finally a manner is presented in which the data and header information might be stored in temporary, disk-oriented files for independent access by the computer

B.2 UNPACKING COST

The unit of cost in this analysis in the CPU second.

Some simple timing tests were run to determine the cost of unpacking data. The specific computer was a CDC Cyber 176 (with a millisecond clock), but the purpose of the exercise was to demonstrate relative values, not absolute ones.

The basic Fortran construction was the following.

```
Do   10  I = 1,1000
J = J+1
10 CONTINUE
```

In several repetitions, the total CPU time for this was always less than one millisecond. The constructions listed in Table B.1 were then run (with the indicated elapsed CPU time for the entire construction, which includes the above simple FORTRAN loop).

TABLE B.1

No. Elements in Loop	Construction	Loop CPU Time
1000	Unformatted (binary) reads of 512 words (3840 bytes)	0.610
1000	Extractions of a 17 bit field-from a constantly shifting area which did not respect word boundaries.	0.006
1000	Extractions of a 9 bit field- from a constantly shifting area which did not respect word boundaries.	0.005

The field extractions were performed by a small assembly language routine, but everything else was pure FORTRAN. The quoted CPU times are literally CYBER-176 times (seconds), but the only relevant point (of general applicability) is the ratio of times. Typical in-core operations required to implement the unpacking are at worst only one percent of I/O time - and other measures would look much better. It is obvious that a tremendous payoff is available in the data compaction area (with the accompanying reduction in I/O).

The above arguments are aimed at relevant considerations for any archive format. They do not apply to exchange formats (which must be character oriented for

compatibility) or in-house formats while the data set is being worked on, which probably should be unpacked, unbiased and unscaled for ease of use in processing codes.

B.3 THE DELTA FORMAT

This section illustrates practices to save storage space on tape. For the purposes of illustration assume an original format of 4 bytes (of 8 bits each) per variable in an IBM floating point format.

B.3.1 Even with the limited precision of the original format it is far more than is needed to store oceanographic parameters which typically require accuracies of one part in ten thousand. The extra space devoted to the floating point exponent is hardly used at all.

B.3.2 Suppose we scale and bias things such that everything can be represented by a positive integer. The bias and scale factors must be stored of course, but only once per station. Suppose we make the further assumption that five significant figures are totally adequate - certainly not unreasonable for physical measurements at sea. Now, each datum can be represented in only 17 bits:

$$\text{datum} \leq 99999 < 131072 = 2^{17}$$

Our requirements compared to the original are now

$$\frac{17 \text{ bits}}{4 \text{ bytes} \times 8 \text{ bits/byte}} = 53\ \%$$

We have certainly gained something, but at what cost? In order to use our scaled, biased values packed away on tape, we must unpack bit fields and apply the inverse scale and biassing factors. It has been shown that this cost is negligible compared to the I/O costs.

B.3.3 Let us make another somewhat radical change in the storage mode. In B.3.2 each value is allocated a full seventeen bits. If we adopt a "delta" format, we can make another dramatic savings. Represent the series as a base value (the first one, for example) plus a series of increments between consecutive values. For example, the series (125, 126, 128, 129, 132, 130, 131, 133) would be represented as a base value of 125 plus a delta-series of (1, 2, 1, 3, -2, 1, 2). If we allow a full 17 bits for the base value and 9 bits for each delta value, we see that the space required goes (for N values) from 17 N bits to 17 + 9(N-1) bits. The former is exactly 17 bits per datum, while the latter coverages out to about 9 bits per datum. These 9 bits can handle deltas of ±256 parts out of the 10,000 in the dynamic range (i.e., ±2.5 %). As before, the additional computer processing needed to recover the data to usable format is negligible. We now require only approximately (9/32) of the original storage space. This is a net saving in space of over 70 percent.

As indicated, a 9 bit delta format is reasonable for oceanic changes of ±2.5% of the dynamic range. Some stations may require more than this, some less. With additional bookkeeping, like a field that indicates the number of bits per delta, one can avoid making the choice between losing data and being forced to define a delta field so large as to provide too little savings.

B.4 TOTAL STORAGE REQUIREMENTS

This section is concerned with the magnitude of the problem - simply how much data there is to store. Molinelli and Kirwan (1980) estimate that there are 150,000 STD stations presently in hand and 100,000 CTD stations. The present growth rates of relevant factors indicates that the total inventory could triple in the next five years. Assume that a typical data cycle consists of measurements of pressure, temperature and salinity, and further that there are typically 1,000 data cycles per station (every decibar to a pressure of 1,000 decibars). This yields a total of 7.5×10^8 data values in the present inventory. Space for header information, counters and narrative commentary must be added to this, but the additional information is less than the basic data requirements. Also, different techniques for data storage will have less effect on the header portion of the station which contains a great deal of alphabetic text.

To estimate the number of tapes required to store all these stations the following calculation is performed. The sample calculation below makes specific reference to IBM concepts, but has general application as far as the bottom line (maximum information stored on a tape) is concerned.

Maximum IBM tape record size ≈ 32750 bytes

$$\frac{32750 \frac{\text{bytes}}{\text{record}} \times 8 \frac{\text{bits}}{\text{byte}}}{1600 \frac{\text{bytes}}{\text{inch}} \times 8 \frac{\text{bits}}{\text{byte}}} \qquad \text{20.5 inches per record of 32750 bytes}$$

Allowing 0.5 inches for an inter-record gap:

$$\frac{2400 \frac{\text{feet}}{\text{tape}} \times 12 \frac{\text{inches}}{\text{foot}}}{21 \frac{\text{inches}}{\text{record}}} \approx 1370 \text{ records per tape}$$

Finally: $$1370 \frac{\text{records}}{\text{tape}} \times 32750 \frac{\text{bytes}}{\text{record}} \approx 45 \times 10^6 \frac{\text{bytes}}{\text{tape}} .$$

(remember, these are 8-bit bytes)

If we assume (for the sake of numbers) that each of the 7.5×10^8 data words in the "original" format requires 4 bytes of storage, we are led to the following:

$$\frac{7.5 \times 10^8 \text{ datum} \times 4 \frac{\text{bytes}}{\text{datum}}}{45 \times 10^6 \frac{\text{bytes}}{\text{tape}}} = \frac{30 \times 10^8}{45 \times 10^6} \approx 70 \text{ tapes with headers.}$$

This assumes essentially one hundred percent efficiency - seldom realized, and is only for the present inventory, without tape backups (always a prudent policy). Header information and less than 100% efficient use of tape space could nearly double the tape requirements. The 70% savings offered by the delta format represents a savings of 50 tapes, of the 70 for data storage. Assuming another 50 tapes are required to handle header information and unused lengths, one is comparing 70 tapes using the delta format to 120 tapes without. This amounts to over a 40% reduction in total tape demands by going to a delta format. Word packing of non-character header information can further the savings.

B.5 STORAGE REQUIREMENTS FOR STATION HEADER INFORMATION

All the information about the station and cruise provided by the exchange tape must be stored on the permanent tape. The information need not be stored in the same format, so adjustments for the sake of saving space should be made.

In the exchange format, the station and cruise information include the time and space coordinates and cruise/station identifiers in the cruise header and station header records. Also included are the surface and classical observations, the data cycle definitions and the text describing collection and processing techniques. In the character format used on the exchange tapes, this information requires a total number of bytes (TB) that is a function of the number of levels of classical observations (NCL), the number of classical parameters at each level (NCP) and the number of STD/CTD parameters at each level, i.e., data cycle (NSP). The form is

$$TB = 1414 + NCL * 38 + NCP * 463 + NSP * 523.$$

For a typical station, classical measurements might be made at 10 levels for each of 4 parameters (pressure, temperature, salinity and dissolved oxygen) while STD measurements are made of 3 parameters (pressure, temperature and salinity). In this case the number of characters required to store the cruise-station header information is 5,215 bytes.

More than half of these bytes are text, the remaining bytes are decimal digit numerals which can be

represented by binary integers and packed across word boundaries. Each decimal digit requires no more than 4 bits which is half of an 8 bit byte (the unit being used for all these discussions of storage requirements). Therefore, if numerals are packed, only half as many bytes are needed to store them. Delta formatting is generally not appropriate for header parameters. The split between total text bytes (TTB) and total numerical bytes (TNB) is given by the following relations:

$$TB = TTB + TNB$$

$$TTB = 866 + NCP * 334 + NSP * 361$$

$$TNB = 548 + NCL * 38 + NCP * 129 + NSP * 162.$$

After packing, the new number of bytes required (TB') is given by:

$$TB' < TTB + 1/2 * TNB.$$

So for the same typical station as above the storage requirements are now less than 4,250 bytes.

More savings are possible when one realizes that a great deal of this information is constant for all stations on a cruise. Since the data are being stored in cruise order this cruise information can be stored at the beginning of the cruise and only those elements that change need be stored with the station data. Retrieval is made simple if text and numbers are separated. The sequence on the permanent storage tape would be as shown in Figure B.1.

SEQUENCE OF INFORMATION IN CRUISE FILE
ON PERMANENT STORAGE TAPE

CRUISE DATA
including time and place and collection and processing numbers (packed)

CRUISE TEXT
including cruise identifiers and collection and processing text

STATION #1 HEADER DATA
including time and place and classical observations and adjustments to the cruise reported collection and processing numbers (packed)

STATION #1 HEADER TEXT
including station identifiers and adjustments to the cruise reported collection and processing text

STATION #1 DATA
includes data cycles (packed, delta format)

STATION #2 HEADER DATA
...

etc.

Figure B.1

The above description identifies the approach sufficiently for this report. Further details need not be developed until the stage at which the approach is to be implemented.

B.6 TEMPORARY STORAGE REQUIREMENTS

Once the data have been transferred to core by unpacking, unbiasing and unscaling, further translations should be avoided. After all operations on and/or modifications to the data set have been made, then it should be transferred back to tape. (Files not affected by modifications but that share a tape with files that are, should be transferred to the new tape, whenever possible, using computer utility routines rather than a read tape, write tape sequence of FORTRAN commands).

The data must be on line, i.e., available to the computer by disk I/O commands, while programs are executing that use that data. The data should be available one station at a time so a random access mode of storage is required. This implies using a fixed record size which can be difficult to do when station sizes vary so dramatically. A solution is achieved by selecting a reasonable record size (say 1 disk track $\approx 5 \times 10^4$ bytes) and allowing a station to occupy as many random access records as needed. The first record can contain the total number of records required and each record can contain the disk record ID of the next record. The data retrieval routine will then use this information to reassemble an entire station in core. At the end of the day the disc file can be transferred to a tape in disc image using utility programs. The disk then may be reloaded from this tape the next day, or whenever work on the data will recommence.

A catalogue of which data are in disk image on which tape or disc must be kept. The catalogue must also indicate the first disk record for each station in the data set.

The 5 x 10^4 bytes on each track are sufficient room for the header information (5 x 10^3 bytes for a typical station, see B.5) and over 3500 data cycles (for the three parameter cycle of the typical station, allowing 4 bytes per floating point parameter). With 3000 to 5000 tracks readily available (D. Knoll, NODC, personal communication) the station capacity of the system is sufficient for most purposes.

Unlike the permanent storage format, the cruise fixed information should not be separated from the station information because from the temporary storage the stations must be available independently. This is true for the production of geographically sorted copies of the data for dissemination and for the performance of quality control tests.

LME